SOCIAL FORCES

AN INTERNATIONAL JOURNAL OF SOCIAL RESEARCH
ASSOCIATED WITH THE SOUTHERN SOCIOLOGICAL SOCIETY

VOLUME 74: NUMBER 1 SEPTEMBER 1995

IN GENERAL

1. Mail manuscripts (5 copies) to: Social Forces, Department of Sociology, Hamilton Hall, University of North Carolina, Chapel Hill, NC 27599-3210.
2. The paper must not be submitted to another journal while under review by SF.
3. Enclose a $15 processing fee. Papers are not reviewed until this fee is paid. The fee is waived for papers solely by currently enrolled students.
4. Manuscripts are not returned to authors.
5. Authors transfer copyright to SF. Permission fees for reprint of articles are split between authors and journal.

PREPARATION OF MANUSCRIPTS

1. For review, papers may be submitted in the form used by SF, American Sociological Association or other major sociological journals. Authors of accepted manuscripts are responsible for putting them into SF form.
2. **Format, Organization, and Style** Submit 5 clearly legible copies on 8 ½ by 11 good quality white paper. Double space typing throughout. For an accepted paper being revised to fit SF style, leave at least 1" margins; do not justify margins if using a word processor.

i. **Order** Organize the manuscript in this order: cover page; abstract; text; endnotes; references; tables; figures.

ii. **Cover Page** Give title; author(s); affiliation(s); and a footnote (*) indicating name and address of the author to whom requests for offprints or other correspondence should be sent ("Direct correspondence to ___") and acknowledgment (if any) of financial or other assistance.

iii. **Abstract** On a separate page, preceding the text, write a summary, 125 or fewer words (70 or fewer for a Research Note).

iv. **Endnotes** Use only for substantive comments, bearing on content. Number consecutively from 1, double space, and append on a separate page.

v. **Tables** Type each table on a separate page. Insert a location note — "Table 1 about here" — in the text. For substantive footnotes relating to a table or figure use lowercase letters. Use tabs to separate columns.

vi. **Figures** Must be artist-drawn, camera ready. Do not send original copy with a manuscript submitted for review. For review, figures may be roughly drawn but they must be neat and legible. For footnotes to figures, use same symbols as for footnotes to tables.

vii. **Mathematical Symbols** (and others likely to be obscure to compositor) Clarify with encircled words written in the margin.

viii. **References in Text** Indicate sources as illustrated below:

1. when author's name is in text — Lipset (1960); when author's name is not in text (Lipset 1960)
2. use page numbers only for direct quotations or specific notes or table — (Braudel 1969:213-14)
3. for more than 3 authors use "et al."
4. with more than 1 reference to an author in the same year, distinguish them by the use of letters (a,b,c) with year of publication (1975a)
5. earlier publication should precede later publication in brackets with parentheses (Tocqueville [1835] 1956)
6. enclose a series of references — in alphabetical order — in parentheses, separated by semicolons (e.g., Adler 1975; Adler & Simon 1979; Anderson, Chiricos & Waldo 1977; Bernstein et al. 1977; Chesney-Ling 1973a, 1973b).

ix. **References Following Endnotes** List authors alphabetically, by surname. Spell out first names of all authors and editors. For authors with more than one work cited, list works earliest to latest. For articles, next give title of article (caps and lower case), name of journal, volume number, and pagination. For books and monographs, give title, followed by publisher.

x. **Format of References**
Please spell out the first names of all authors and editors, unless they use only their initials or a first initial and a middle name in the source cited (e.g., Paul Radin, T.S. Eliot, and J. Owen Dorsey).

Elder, Glen H. 1975. "Age Differentiation and the Life Course." Pp. 165-90 in *Annual Review of Sociology*. Vol. 1, edited by Alex Inkeles, James Coleman, and Neil Smelser. Annual Reviews.

Myrdal, Gunnar. [1944] 1962. *An American Dilemma*. Harper & Row.

Ritzer, George. 1975a. *Sociology: A Multiple Paradigm Science*. Allyn & Bacon.

____. 1975b. "Sociology: A Multiple Paradigm Science." *American Sociologist* 10:156-67.

Computer Files Authors analyzing data available in this format should cite their source, indicating producers and distributors as in these examples:

Louis Harris and Associates. 1975. Harris 1975 Nuclear Power Survey #2515 [computer file]. New York: Louis Harris and Associates [producer]. Chapel Hill: Louis Harris Data Center, University of North Carolina [distributor].

U.S. Bureau of the Census. 1979. 1970 Census of Population and Housing, Fourth Count Population Summary Tape [computer file]. Washington: U.S. Bureau of the Census [producer]. Rosslyn, Va.: DUALabs [distributor].

SOCIAL FORCES

SEPTEMBER 1995, VOLUME 74, NUMBER 1

ISSN 0037-7732 Library of Congress Catalog Number 24-31023

Second class postage paid at Chapel Hill, N.C. 27514 and additional mailing offices.
Postmaster: Send 3579 to Social Forces, UNC Press, P.O. Box 2288, Chapel Hill, NC 27515

Claims for undelivered copies must be made within the month following the regular month of publication. The publisher will supply missing copies when losses have been sustained in transit and when the reserve stock will permit.

Social Forces is issued quarterly:
September, December, March, June.

Social *Forces* is published by the University of North Carolina Press for the Department of Sociology, with the assistance of the Institute for Research in Social Science, University of North Carolina at Chapel Hill.

Rev. 7/3/95

BOOK REVIEWS

The Practical Consequences of Sociology's Pursuit of "Justice for All"*

THOMAS C. HOOD, *University of Tennessee, Knoxville*

Abstract

The following essay explores the consequences of American sociology's commitment to the value, "justice for all." Concentrating on the idea of substantive justice, the article assesses consequences for the generation of sociological knowledge, the teaching of sociology, and sociology's place in the development of public policy. Several conclusions are drawn about the nature of justice and the practice of sociology. Sociological knowledge is viewed as particularistic. The teaching of sociology competes with other disciplines in curriculum politics in American higher education. Sociology, properly taught, has much to offer; its contribution to public policy suggests a strong respect for other value systems and a recognition of global interdependence.

Thirty-three years ago, Pitirim A. Sorokin's (1962) article, "The Practical Influence of 'Impractical' Generalizing Sociological Theories," was published. His impressive, and perhaps self-serving, documentation of the impact of the theories of Marx and Gandhi, Aristotle and St. Augustine, Hegel, Marx, Spencer, and Freud serve to support his conclusion that the role of builder of grand systems of social thought is a most important mission. Kurt Lewin has been quoted as saying that there is nothing as practical as a good theory (Marrow 1969). This article will not develop a new or groundbreaking definition of *justice.* The arguments offered will not solidify sociology's claim to the position of queen of the sciences. The following essay will develop the following thesis. *American sociology has a basic value commitment and advocacy stance embedded in its theory and practice. This value is that justice practiced correctly is blind — all are equal before the law. All persons are entitled to a chance to succeed. This commitment leads to a critical stance toward social institutions that are seen as organizations and*

** This article is a slight revision of the presidential address given 7 April, 1995 at the Southern Sociological Society, Atlanta. The author wishes to thank Diana Harris, Donald Clelland and Michael Benson for criticism and suggestions and to express particular gratitude to Bennett Judkins and John Gaventa for the inspiration provided by their scholarly work and their work in organizing the 1995 program for the Southern Sociological Society Annual Meeting.* *Direct correspondence to the author at Department of Sociology, 901 McClung Tower, University of Tennessee, Knoxville, TN 37996-0490. E-mail (Internet): tomhood@utkvx.utk.edu.*

 Social Forces, September 1995, 74(1):1-14

customs that tend to perpetuate inequalities of economic position, prestige and political power rather than working to eliminate them. After stating three points regarding the definition of justice, the essay develops the consequences of this value position for sociological knowledge, for sociology as a teaching enterprise in the American college and university, and for sociology as it has implications for social policy.

Defining Justice

Definitions of *justice* abound. An excellent review article from a social psychological point of view appears in the new volume, *Sociological Perspectives on Social Psychology*, edited by Karen S. Cook, Gary Alan Fine, and James S. House (Hegtvedt & Markovsky 1995). Their article concentrates on research on distributive justice and procedural justice.[1] This article concentrates on what the authors of *Habits of the Heart* call substantive justice or "a matter of the institutional order of society as a whole and its justice or fairness" (Bellah et al. 1985:334).

Justice requires agreement about the purposes or ends of life and not simply procedures for resolving disputes or the fairness of the society's system of rewards or of its distribution of goods and opportunities.

I agree with the authors of *Habits of the Heart* (Bellah et al. 1985) and *The Good Society* (Bellah et al. 1991) when they say that the conception of justice in American society is founded in the the Judaeo-Christian tradition. Whether talking about the first governor of Massachusetts, John Winthrop, Thomas Jefferson, or a contemporary Jewish child, the subjectivity of American justice is legitimated by belief in some higher authority. Consider the Exodus story.

> [It] binds together all sorts of Americans from the beginnings of our history to the present. In 1988 the thirteen-year-old daughter of a Jewish immigrant from South Africa offered this text as the reading for her bat mitzvah: "When you reap your harvest in the field, and forget a sheaf in the field . . . it is to go to the resident alien, the orphan and the widow, that the Lord your God may bless you in all your enterprises. . . . You must remember that you were once a slave yourself in the land of Egypt; that is why I am commanding you to do this" (Deuteronomy 24:19). She is retelling a fragment of a story that Americans from John Winthrop to Martin Luther King have told over and over again to make the same point. It is a story that helps to define the consciousness of our society. A good society for Americans is one in which our children may still retell such a story with hope that they can at least partially live up to its meaning. As long as we define our national identity in terms of being able to create and sustain such a society, we will have a self-understanding that can be defended and pursued in the great society of our interdependent world. (Bellah et al. 1991:247)

Two points to remember about justice as used in this essay are that the concern is substantive justice and that in American sociology the foundation for understanding substantive justice comes from the Judaeo-Christian tradition. The third point is that American sociology often selects problems to investigate that examine the American ideal of "justice for all." This is a crucial point in that the *all* in sociology includes all subcultures and all societies. The cultural

relativism advocated by Ruth Benedict ([1934] 1959) in the 1930s has become the accepted value position of the discipline:

> We must be willing to take account of changing normalities even when the question is of the morality in which we were bred. Just as we are handicapped in dealing with ethical problems so long as we hold to an absolute definition of morality, so we are handicapped in dealing with human society so long as we identify our local normalities with the inevitable necessities of existence. (234)

Sociologists have responded to her call for this commitment and cultural relativity (now called "multiculturalism") is taught not only in sociology courses but across the curriculum. With Benedict a number of people would say that we are well on the way to accomplishing the task of formulating "a more realistic social faith accepting as grounds of hope and as new bases for tolerance the coexisting and equally valid patterns of life which mankind has created for itself from the raw materials of existence" (240).

I suspect that attentive readers are thinking, "Didn't he just claim that American conceptions of justice are founded in the Judaeo-Christian ethic, which is a particular world view and value system!" "Aren't sociologists committed to the view that *all* value systems are legimate, have their own logic, and deserve respect. Isn't there some potential for conflict here?" The answer is Yes. There is potential for conflict, but these are the three definitional starting points for the essay.

1. Substantive justice is a value embedded in the fairness of the institutional order of the society.

2. The American conception of fairness or what is just is derived from or legitimated by a belief in a higher conception of justice or righteousness found in the Judaeo-Christian tradition.

3. Justice for all as taught by sociologists means that justice should respect the value system of the particular subculture or society within which the action is seen as just or unjust. (Thou shalt not violate the "prime directive" of imposing one's value judgments on a subculture or society different from one's own.)

Justice and the Creation of Sociological Knowledge

Given this conception of justice what is the current impact on the practice of sociological research? Over the last 15 years or so, American sociology has emerged from a period during which logical positivism and sample surveys using quantitative operational definitions were the dominant form of data collection. Now qualitative methodologies and alternative epistemologies are being advocated as a way to achieve a fuller understanding of human social behavior. The early works, Monica B. Morris, *An Excursion into Creative Sociology* (1977) and Jack D. Douglas, *Introduction to the Sociologies of Everyday Life* (1980), surveyed the emerging movements of ethnomethodology, phenomenology, existential sociology, trends in symbolic interactionism and the work of Erving Goffman. Increasingly people came to realize the importance of language not just in the sense of precise operational definitions but in the contribution of language and nonverbal communication to the expectations and understandings

that constitute microfoundations of social order. In some ways this was a natural response to the philosophical treatise published by Berger and Luckmann in the mid-1960s, *The Social Construction of Reality* (1966).

In addition to this trend, by the mid-1970s, people were criticizing otherwise fine pieces of research for the implicit value judgments in the descriptive statements about the populations they were studying. For example, some reviews of Kai Erikson's, *Everything in its Path,* (a Sorokin Award winning book) criticized the description of the inhabitants of Buffalo Creek and other mountain communities as biased (Billings & Maggard 1978; Dynes 1978). Other works, such as Rodger Cunningham's *Apples on the Flood* (1987), critique the work of Weller and other scholars on Appalachia by questioning their failure to understand the point of view (subjective meanings) of those they were studying.

The newer methods listed above ask the question that Benedict said we must ask about any society or subculture — namely, what is the system of values in that society or subculture. Cunningham shows the failure of many scholars to capture the essence of Appalachians. He suggests that these scholars, like Arnold Toynbee, Michael Harrington, or Jack Weller, may approach the Appalachian with the ideal of objective observation, but use a standard of the outside culture as a reference point and view the Appalachian as a child not capable of acting responsibly in its own best self-interest. The dominant culture "America" must tell "hillbillies" how to act responsibly. Recent shows like "The Beverly Hillbillies" do not offend our sensibilities in the way that "Amos and Andy" did, because we regard the caricatures offered as nothing more than exaggerated portrayals of real differences.

Cunningham (1987) argues that:

> the dominator implicitly denies the dominated an inner self and an ability to think and act on one's own behalf (Laing 1971:101). Such disconfirmation can take "benevolent," patronizing forms, which are perhaps the most damaging — for example, the attitude expressed in Jack Weller's (1966) injunction, "Let us begin to guide the mountain man toward wholeness" (150). The attitude behind such a statement was laid bare by Rupert Vance when he wrote that Weller, "because he came as a missionary . . . brought the objectivity of the stranger. Finally he came to know these people better than they know themselves" (v). Here the positivist ideal of "objectivity" is explicitly invoked as a means of gaining knowledge of the other which is superior to one's own knowledge of who he or she is. Truly, "violence cannot be seen through the sights of positivism" (Laing 1967:51). To connect this with our earlier themes, let us note that addressed to an adult, the call to "grow up" (in social terms, to attain "wholeness") is by implication just as infantilizing as the command to remain a child (Rogin 1975:210). Just so the call to "us" outsiders to "guide the mountain man toward wholeness" disconfirms the real agency of the mountaineer, refuses to recognize his or her real existence, and thus defeats its stated goal. (118)

Slim's Table, a recent study of African American males, shows the promise for community in sustaining American culture (Duneier 1992). Carefully constructed and based on participant observation, the work demonstrates the failure of past studies of this group to avoid stereotypes commonly voiced in the press and even sociological literature. Duneier points out that past ethnographies and studies of black males by American sociologists have many positive qualities. Nevertheless, the authors made "patently inaccurate inferenc-

es about blacks or unfair generalizations from selective samples that reinforce, rather than question, the most basic stereotypes. . . . In the process of developing deeper understandings, sociologists casually make demeaning assertions about the nature of black men" (142).

In one of the most perceptive treatments of this problem, Patricia Hill Collins (1990) points out that positivist methodological approaches require the distancing of the researcher from the object of study; removing emotions from the research process; limiting the role of ethics and values in the research process to the rights of human subjects; presenting results in an adversarial fashion to determine which argument "can withstand the greatest assault and survive intact" (205).

We are indebted to a number of writers for suggesting to us that alternatives to positivism, which may be quite appropriate as an epistemology for investigating universalistic relationships, exist. Collins has built on the work of Gilligan (1982) and on *Women's Ways of Knowing* (Belenky et al. 1986) to suggest one alternative epistemology to the assumptions of positivism. Patricia Hill Collins's award winning book, *Black Feminist Thought,* offers the reader an alternative epistemology for exploring the kinds of social relationships that Parsons and Shils (1962) labelled "particularistic." In her chapter, "Toward an Afrocentric Feminist Epistemology," she offers the following dimensions of an alternative epistemology. First, she presents an emphasis on concrete experience with "practical images as its symbolic vehicles." Second, she provides dialogue and connectedness as criteria for methodological adequacy. "A primary epistemological assumption underlying the use of dialogue in assessing knowledge claims is that connectedness rather than separation is an essential component of the knowledge validation process" (Belenky et al. 1986:18). Third, an ethic of caring is posited which includes the idea of unique individual expressiveness, the use of emotion in dialogue, and the ability to demonstrate empathy. Finally, the epistemology includes an ethic of personal accountability (Collins 1990:201-20).

Assessments of an individual's knowledge claims simultaneously evaluate an individual's character, values, and ethics. African Americans reject the Eurocentric, masculinist belief that probing into an individual's personal viewpoint is outside the boundaries of discussion. Rather, all views expressed and actions taken are thought to derive from a central set of core beliefs that cannot be other than personal (Collins 1990:218).

Obviously this kind of epistemology takes a very different kind of view of the statement with "justice for all" than does a positivistic approach which emphasizes rational, value-free, impersonal, and "disinterested" (Merton 1973:275-77) discussion to arrive at truth. If ways of knowing are based within a group or within the capacities generated through individual experiences, certain topics will be distorted by survey research that assumes that individual biases of question interpretation are distributed randomly through the population being surveyed.

So what does sociological knowledge look like? First of all, sociological knowledge is not value-free but conditioned by the value system in place within the various social orders that sociologists investigate. We contend that investigators are limited by the value system that they endorse and by their ability to

understand and interpret value systems different than their own. In a sentence, sociological knowledge is embedded in the investigator's value system due to the difficulty in separating observed social action into descriptive, emotional and evaluative components.

Second, all sociological knowledge is to use a word, "particularistic." One of Talcott Parsons pattern variables (1962) reflected on how normative expectations for behavior were either applied universally ("justice is blind") or particularistically ("I can make an exception for you because you are my friend, my sister, my brother, or because you are rich"). This idea that normative expectations are not universally applied in the course of everyday life has been widely supported by experimental and survey research (e.g., see Stouffer & Toby 1951). This simply means that when we look at day-to-day conduct of social life, people are constantly making allowances with regard to the normative expectations that they have depending on the prior relationships that they have had with the individual or individuals in question. The individual or individuals in question in turn enact behavior based upon their knowledge of the cultural background in which they are interacting. If that knowledge is limited, their behavior may appear strange. I recently attended a symphony concert. Attending the same concert was a group of men dressed as women. Their behavior and dress was appropriate for middle-class women attending a somewhat dressy occasion, but clearly not all of them recognized or knew the norm of waiting until the end of all the movements of a particular piece of music before breaking into applause for the performance. These men had learned appropriate dress and demeanor for the opposite gender but some had not learned appropriate behavior for showing appreciation at a symphony concert. The behavior that we enact is always based on experience conditioned by the relationships that we are trying to maintain. This behavior emerges against the background of the universalistic understandings that exist within our culture.

What are these universalistic understanding or expectations about life within our culture? Sumner ([1906] 1960) called them mores. Certainly they are not represented by the laws that our governments enact. Those social products are no more than aging photographs of the state of the moral order or consensus at a particular point in time. The universal expectations about behavior are like Plato's forms; we glimpse them only as shadows thrown up on the walls of our little caves.[2] Yet we strive to understand the universal forms from an investigation of the particular because we wish to live orderly lives. Human beings want to be able to predict the behavior of other human beings at least to the point that our daily lives can coordinate to some degree. As Kingsley Davis (1949) said many years ago, consistency or predictability in behavior is the main criterion we use for normality or sanity.

Justice and the Place of Sociology in the American College

Since sociology is not a value-neutral pursuit and since the politics of American higher education are anything but objective, sociology can easily become an area of trouble, controversy, and conflict.

Although the percentage has declined to around 70%, according to Horowitz (1993) in the last 25 years, most sociologists remain employed in academic settings. Many of those employed in academic settings have teaching appointments or have duties that include teaching, research, and administration. Because our field includes the study of formal organizations, it seems appropriate that we have engaged in critical studies of higher education. Noble's *America by Design* (1977) reveals how far the modern American university and some colleges have come from the medieval concept of the ivory tower. The indebtedness of budgets to external support by business, industry, and governmental programs raises questions about the function of the pursuit of knowledge within the university. And because sociologists are taught to raise questions about the impact of such interdependent relationships, sociology departments have been viewed as troublesome and troubling for both sociology and the university itself. *For sociology the commitment to justice for all as a basic value in sociological investigation threatens some of the long standing institutional ties of the university to lawyers, priests, physicians, and the business community.* As CBS's "60 Minutes" recently noted some institutions derive half their funding from grants and contracts for research sponsored by government or the private sector.

Bourdieu (1993) in the book, *Sociology in Question,* puts this nicely:

> The type of social science that one can do depends on the relationship one has to the social world, and therefore on the position one occupies within that world.
>
> More precisely, this relation to the world is translated into the function that the researcher consciously or unconsciously assigns to his practice and that governs his research strategies — the objects chosen, the methods used, and so on. You may make it your goal to understand the social world, in the sense of understanding for understanding's sake. Or you may seek techniques that make it possible to manipulate it in which case you place sociology in the service of the management of the established order. . . .[3]
>
> A good number of those who describe themselves as sociologists or economists are social engineers whose function is to supply recipes to the leaders of private companies and government departments. They offer a rationalization of the practical or semi-theoretical understanding that the members of the dominant class have of the social world. (13)

A second area of conflict in higher education is the competition for students. Enrollments in the contemporary university are driven more by what is required to earn a specific degree than by the attraction to a particularly brilliant lecturer or instructor. The bachelor of arts and bachelor of science degrees are awarded on the basis of the accumulation of satisfactory grades in a specified set of courses. The political debate in the university community arises over which courses will be listed as fulfilling general education requirements and which courses will be required for the completion of a major and what majors should be offered. The logic here is obvious. The more programs in which a given course is listed for credit the larger number of students who will potentially enroll in the course when it is offered. The larger the enrollment the more frequently the course can be offered to meet the demand. The more sections of a course needed to meet the demand the more staff needed to teach the course regularly. The more staff needed to teach the course regularly the easier it is to justify the request for additional permanent staff. The larger the permanent staff the more diversity possible in representing different areas in the

discipline or the easier it is to have more than one specialist in a given area in the discipline making more community and collegial exchange possible. In order to profit from a curriculum revision, a sociology department would probably have to form an alliance with other departments who would boost the cause of sociology courses in return for sociology faculty lobbying and for votes for the allied department's courses.

In the last major curriculum revision at the University of Tennessee, two years (or intermediate competence) in a foreign language became a requirement of all degree programs in the College of Arts and Sciences. For several years this change created a staffing boom in terms of new positions in the language departments at the school. A similar thing happened in the English department where a year of English composition became a university-wide degree requirement. What is "just" here is determined by which majority prevails in the revision of general education and requirements for majors. Justice for all in the sense of an equal opportunity to attract students is not likely to have much of a following in curriculum debates.

Sociology does have a unique contribution to make to general education. Few people seem to be commenting on this fact today.

One of the most eloquent statements about the contributions of sociology to a general education has been made by Robert Bierstedt (1964). He details seven ways in which sociology as a discipline contributes to the making of an educated person:

> Like history in particular, the study of sociology liberates the student from the provincialisms of time, place, and circumstance and frees him from the constrictions of his natal culture. Secondly, it introduces him to the role of logic and of scientific method in the acquisition of knowledge and thus contributes thirdly, to this sense of order and to his methodological sophistication. In the fourth place, sociology is a discipline that spans two cultures, the scientific and the humanistic, using as it does the method of science to explore the concerns and affairs of humanity. In the next instance, the fifth, . . . sociology initiates and keeps at the front of student awareness the ancient problem of the relationship between society and the individual (1964:54). As a sixth point . . . sociology is not content with description, however thorough it may be, or with analysis, however profound. It has a more fundamental quest. It seeks also the causes of things — things as they are and as they have come to be. (51)

Finally Bierstedt argues that (with some unfortunate exceptions) sociologists write with good and sometimes striking literary quality.

Space will not permit elaboration of these points. Throughout my career I have heard many discussions of sociology and sociologists from those outside the field condemning us for poor scientific method, for poor writing, for provincial and biased teaching. These criticisms are not supported by the evidence. But our silence in the face of the criticism expresses our tacit agreement that our colleagues do not do the best they know. The critics receive our support whenever we tolerate less than the best writing and most careful investigation in ourselves and in our students. Further, the attacks on sociology are justified when we too easily succumb to accepted stereotypes and generalizations rather than challenging them.

Teaching sociology demands our best. Doing research in sociology demands our best. Because we are participants in the societies and cultures that we study

we are constantly in danger of importing the values and the perspectives of that society and the subcultures in which we participate into our teaching and research. Sociology is not chemistry. We often interact with those we study and whose behavior we teach. Even when we do not interact we bring opinions about them into our research and teaching about these extraordinary groups. Those in other disciplines recognize this danger; they aren't stupid. They recognize that sociology walks a narrow line between informing and advocating. When we advocate to the point that we only narrowly inform — we cross the line from being educators to propagandists for the American life or some other way of life, as we see it.

Justice and Sociology's Contribution to America's Social Policy

The above definition of *justice* referred to the widely discussed books, *Habits of the Heart,* and its follow-up volume, *The Good Society* (Bellah et al. 1985, 1991). At the end of the first chapter in *Habits of the Heart,* the authors state the problem of guiding social policy in American society:

> We need to reach common understandings about distributive justice — an appropriate sharing of economic resources — which must in turn be based on conceptions of a substantively just society. Unfortunately, our available moral traditions do not give us nearly as many resources for thinking about distributive justice as about procedural justice and even fewer for thinking about substantive justice. (26)

As noted earlier, Bellah and his team ground their approach to substantive justice in ideas derived from the Judaeo-Christian tradition. Early thinkers such as Jefferson recognized a fundamental contradiction between the value, "justice for all," and American institutions. In Jefferson's inaugural address, he called for "equal and exact justice to all men of whatever state or persuasion, religious or political." While he certainly believed in the procedural justice of our legal system, he could not forget that there is a higher justice that sits in judgment over human justice: "the laws of nature and of nature's God." In considering the continued existence of slavery, Jefferson wrote, "Indeed I tremble for my country when I reflect that God is just; that his justice cannot sleep forever." The profound contradiction of a people fighting for its freedom while subjecting another to slavery was not lost on Jefferson and gave rise to anxiety for our future if this contradiction were not solved (Bellah et al. 1985:31).

In *The Good Society,* the argument is developed that calls Americans to sacrifice, to reform institutions of international finance, to recognize worldwide interdependence (Bellah et al. 1991:243-50).

In an unpublished paper on the topic of environmental justice, I argue that one of the problems with sociological approaches to the environment is a failure to consider the moral significance of the interpretation that past or future generations might place on the use of a particular resource (Hood 1994). Does a particular tree by virtue of the contribution it makes to the beauty of a particular location have a value that transcends the value it would have if cut down and converted into so many board feet of lumber of a particular quality and other wood products? How does an individual compare the aesthetic value and the value in use as lumber? "Justice for all" must include not only those

human beings who are alive now — but, as Edmund Burke would say, those who were living in the past and those who are yet to be born (Burke [1790] 1955:110).[4]

Two other American authors coming from different perspectives have made recent conspicuous pleas for a more inclusive definition of "justice for all." From the political economy perspective, Horowitz in his recent *The Decomposition of Sociology* examines "policy research in a post-sociological environment" (1993:169-82). Horowitz uses the petroleum and energy crises of the 1970s and the Gulf war of the early 1990s, the drug policies and "wars" of the 1980s, and the AIDS epidemic to illustrate issues confronting the U.S. government that break down the familiar and long held pattern of distinguishing between domestic policy and foreign policy. After adding space exploration and environmental considerations to his list, he cites the annual report of the Research Triangle Institute as organized into functional categories rather than five disciplines that have contributed to social research. (Program evaluation research, population and family studies, alcohol and drug abuse, criminal behavior, mental health problems, hazardous-waste management policy, natural-resource damage assessment, risk management, energy resource planning, and market assessment and forecasting are the categories rather than psychology, sociology, economics, political science, and anthropology.)

Horowitz (1993) suggests that:

> We are at the dawn of a new era of unitary policy-making; and we need to reconsider organizational frameworks inherited from a past in which the dichotomy of the national and the international still made some sense. Clearly, the velocity of change, the shrinkage in measuring distances between places, and the shared physical as well as social environment make all past paradigms obsolete. The realities of social research and policy have already transcended old boundaries; only the forms remain to be changed. (178)

Albert Gore Jr. (1993) examines imperatives for international institution building in his examination of the implications of environmental policies in *Earth in the Balance*. Gore speaking as an observer of the world scene suggests that most countries in the world have made three important choices: "first, that democracy will be the preferred form of political organization on this planet; second, that modified free markets will be the preferred form of economic organization; and third, that most individuals now feel themselves to be part of a truly global civilization" (298). Noting the need to recognize the diversity among nation states in any plan to respond to a global environmental crisis, Gore ([1992] 1993) proposes a global Marshall Plan with five major goals:

> 1. Stabilizing of the world population.
>
> 2. The rapid creation and development of environmentally appropriate technologies.
>
> 3. A comprehensive and ubiquitous change in the economic "rules of the road" by which we measure the impact of our decisions on the environment.
>
> 4. The negotiation and approval of a new generation of international agreements.
>
> 5. The establishment of a cooperative plan for educating the world's citizens about our global environment. (305-306)

Gore advocates an overall purpose in establishing the plan of "the establishment, especially in the developing world — of the social and political conditions most conducive to the emergence of sustainable societies — such as social justice (including equitable patterns of land ownership); a commitment to human rights; adequate nutrition, health care, and shelter; high literacy rates; and greater political freedom, participation and accountability" (307).

These thinkers have converged on a central point of policy: (1) American social policy today has international or global consequences; (2) Justice, particularly substantive justice, requires that Americans re-evaluate their position of privilege within the world system. If I read these authors correctly, each of them suggests that we have a need to re-evaluate our own standard of living in terms of the consequences that the maintenance of that standard of living has for the planet. Paul and Anne Ehrlich (1988) compare families in various countries in the world They point out that an average American family affects the environment 40 times more than a family in India and 100 times more than one in Kenya. They continue:

> Conveniences taken for granted by the average American are impossibly beyond the reach of most Kenyan or Indian households — items such as cars, televisions, refrigerators, dishwashers, ranges, air conditioners, and numerous small appliances. The same applies to the dazzling variety of food found in ordinary American supermarkets.
>
> The Brogans' [the name of the American family described] vehicles, two of 180 million cars and trucks in this mobile society, illustrate the culture's demands on resources and the environment. The metals in the vehicles are nonrenewable resources won from their ores with much use of energy and a variety of environmental costs: air pollution emitted at mine and smelter and destruction of ecosystems at the mine site from building roads and dumping overburden. A long list of side effect flows from a vehicle's use: the environmental costs of petroleum extraction, refining, and transport; the pollution and ecosystem destruction from building roads, bridges, parking garages, shopping centers, gas stations; and exhausts that contribute to eye irritation, emphysema, and lung cancer, as well as to acid rain and the greenhouse effect. (938)

Some Personal Conclusions

During my preparation of this article I believe that I learned the following things about *justice*. Whether these have been satisfactorily demonstrated by the excavation of the literature that I have done, I leave to your judgment.

1. Justice is a value widely seen as important in American society and grounded in the Judaeo-Christian religious traditions that have influenced many American organizations and practices.

2. Justice is a property of a relationship.[5] Like love it is particularistic rather than universalistic. Justice is felt; it has both a cognitive and emotional dimension to its existence. This complicates generalizations about justice and raises questions about its epistemological status. This relational quality of justice reveals some of the strengths that newer, more observational methodologies have brought to the discipline of sociology.

3. Justice is practiced within the context of a value system of a particular culture. The notion that justice is blind to individual differences and particular circumstances is directly contrary to the proposition of cultural relativism which suggests that all values

(desires) are learned (embedded) in the value systems of particular societies or subcultures with those societies.

4. Justice requires that we consider the impact of our decisions not just on our own lives but on societies and cultures one third or one half way around the world. Justice is a relational concept and we now understand and affirm the right of not just other species of plants and animals to exist but the right of many cultures to exist. This consideration may mean the creation of new institutions that redefine the meaning of all and the meaning of common good.

In sum my view of American sociology is as a prophetic form of consciousness, struggling for a methodology and strongly committed to the proposition that all human beings are created to live free of social institutions that prevent them from reaching whatever genetic potential they have. I close with a quote from *Slim's Table.* "When we stop trying to feel good about ourselves, or to increase our own power by asserting our innocence, we can begin to look for answers by searching for truth" (Duneier 1992:168).[6]

Notes

1. Hegtvedt and Markovsky's (1995) article provides a concise discussion of various conceptions of justice in the sociological and social psychological literature. Of particular note is the following paragraph:

Social psychological definitions of justice typically refer to particular distribution rules: proportionality (equity), equality of outcomes or opportunity, or distribution according to needs. Cohen and Greenburg (1982) define distributive justice as 'the application of a normative rule to the allocation of resources to recipients' (p. 1). Justice exists when there is congruence between expectations for outcomes based on the normative rule and actual outcomes. What is normative, however, depends on perceptual, interpersonal, and situational factors. Recognition of incongruence between the rule and reality rests on similar factors. Thus, despite explicit definitions such as Cohen and Greenberg's determining what is just remains a relatively subjective process. (259)

2. The very famous passage appears at the beginning of Book VII of *The Republic.* While it is too long to reproduce here, the position revealed should be reviewed as a corrective to the stark empiricism that has dominated much American sociological thought.

3. Bourdieu (1993) offers an example as follows:

A simple example will make this clear: The sociology of religion may amount to research for pastoral purposes that takes as its objects laymen, the social determinants of church-going or abstention; it then becomes a kind of market research making it possible to rationalize sacerdotal strategies for the sale of the goods of 'salvation'. Alternatively it may aim to understand the functioning of the religious field, of which the laity is only one aspect, studying for example the functioning of the Church, the strategies through which it reproduces itself and perpetuates its power — strategies that include sociological studies (initially carried out by a canon). (13)

4. The specific quote from Burke ([1790] 1955) concerns the nature of society:

Society is a permanent contract. . . . It is to be looked on with other reverence, because it is not a partnership in things subservient only to the gross animal existence of a temporary and perishable nature. It is a partnership in all science; a partnership in all art; a partnership in every virtue and in all perfection. As the ends of such a partnership cannot be obtained in many generations, it becames a partnership not only between those who are living, but between those who are living, those who are dead, and those who are to be born. (110)

5. This is exactly the problem. Justice for all implies that each person will be treated as an end and not as a means following Kant's categorical imperative. This means that the distinction between universalism and particularism breaks down and it is difficult for the actor to transcend the limits of a specific relationship.

Parsons and Shils (1962) put it as follows:

The dilemma of transcendence versus immanence. In confronting any situation, the actor faces the dilemma whether to treat the objects in the situation in accordance with a general norm covering *all* objects in that class or whether to treat them in accordance with their standing in some particular relationship to him or his collectivity, independently of the objects' subsumibility under a general norm. This dilemma can be resolved by giving primacy to norms or value standards which are maximally generalized and which have a basis of validity transcending *any* specific system of relationships in which ego is involved, or by giving primacy to value standards which allot priority to standards *integral* to the *particular* relationship system in which the actor is involved with the object. (81-82)

6. The full context of this comment comes on the page before.

Unlike many Americans, the men at Slim's table are beyond believing in the possibility of racial innocence. . . . They see their communities coming apart and believe that more important battles must be waged at this time. They speak candidly about conditions in the ghetto and the wider society. This makes them different from many politicians, journalists, professors, and middle-class blacks and whites who place a higher premium on portraying their own "essential goodness in relation to others" than in searching for truth (Duneier 1992:167).

References

Belenky, Mary Field, Blythe McVicker Clinchy, Nancy Rule Goldberger, and Jill Mattuck Tarule. 1986. *Women's Ways of Knowing*. Basic Books.

Bellah, Robert N., Richard Madsen, William M. Sullivan, Ann Swidler, and Steven M. Tipton. 1985. *Habits of the Heart: Individualism and Commitment in American Life*. University of California Press.

Bellah, Robert N., Richard Madsen, William M. Sullivan, Ann Swidler, and Steven M. Tipton. 1991. *The Good Society*. Knopf.

Benedict, Ruth. [1934] 1959. *Patterns of Culture*. New American Library.

Berger, Peter, and Thomas Luckmann. 1966. *The Social Construction of Reality*. Doubleday.

Bierstedt, Robert. 1964. "Sociology and General Education." Pp. 40-55 in *Sociology and Contemporary Education*, edited by Charles H. Page. Random House.

Billings, Dwight, and Sally Maggard. 1978. "Review of *Everything in Its Path: Destruction of Community in the Buffalo Creek Flood* by Kai Erikson." *Social Forces* 57:722-23.

Bourdieu, Pierre. 1993. *Sociology in Question*. Sage Publications.

Burke, Edmund. [1790] 1955. *Reflections on the Revolution in France*. Bobbs-Merrill.

Cohen, Ronald L., and Jerald Greenberg. 1982. "The Justice Concept in Social Psychology." Pp. 1-42 in *Equity and Justice in Social Behavior* edited by Jerald Greenberg and Ronald L. Cohen. Academic.

Collins, Patricia Hill. 1990. *Black Feminist Thought: Knowledge, Consciousnes, and the Politics of Empowerment*. Harper Collins Academic.

Cunningham, Rodger. 1987. *Apples on the Flood: Minority Discourse and Appalachia*. University of Tennessee Press.

Davis, Kingsley. 1949. *Human Society*. Macmillan.

Douglas, Jack D., Patricia A. Adler, Peter Adler, Andrea Fontana, C. Robert Freeman, and Joseph A. Kotarba. 1980. *Introduction to the Sociologies of Everyday Life*. Allyn & Bacon.

Duneier, Mitchell. 1992. *Slim's Table: Race, Respectability, and Masculinity.* University of Chicago Press.

Dynes, Russell R. 1978. "Review of *Everything in Its Path: Destruction of Community in the Buffalo Creek Flood* by Kai Erikson." *Social Forces* 57:721-22.

Ehrlich, Paul R., and Anne H. Ehrlich. 1988. "Population, Plenty, and Poverty." *National Geographic* 174:914-45.

Gilligan, Carol. 1982. *In a Different Voice.* Harvard University Press.

Gore, Albert, Jr. [1992] 1993. *Earth in the Balance: Ecology and the Human Spirit.* Penguin.

Hegtvedt, Karen A., and Barry Markovsky. 1995. "Justice and Injustice." Pp. 257-80 in *Sociological Perspectives on Social Psychology* edited by Karen S. Cook, Gary Alan Fine, James S. House. Allyn & Bacon.

Hood, Thomas C. 1994. "Environmental Justice in Twentieth Century America: Some Considerations. "Society for Human Ecology 7th International Conference, East Lansing, Michigan.

Horowitz, Irving Louis. 1993. *The Decomposition of Sociology.* Oxford University Press.

Laing, Ronald David. 1967. *The Politics of Experience and the Bird of Paradise.* Penguin.

____. [1969] 1971. *Self and Others. 2d ed.* Penguin.

Marrow, Alfred J. 1969. *The Practical Theorist: The Life and Work of Kurt Lewin.* Basic Books.

Merton, Robert K. 1973. *The Sociology of Science: Theoretical and Empirical Investigations* edited and with an Introduction by Norman W. Storer. The University of Chicago Press.

Morris, Monica B. 1977. *An Excursion into Creative Sociology.* Columbia University Press.

Noble, David F. 1977. *America by Design: Science, Technology, and the Rise of Corporate Capitalism.* Knopf.

Parsons, Talcott, and Edward A. Shils with the assistance of James Olds. [1951] 1962. "Values, Motives, and Systems of Action." Pp. 47-275 in *Toward a General Theory of Action,* edited by Talcott Parsons and Edward A. Shils. Harper & Row.

Plato. 1956. *Great Dialogues of Plato.* Trans. by W.H.D. Rouse. New American Library.

Rogin, Michael Paul. 1975. *Fathers and Children: Andrew Jackson and the Subjugation of the American Indian.* Knopf.

Sorokin, Pitirim A. 1963. "Practical Influence of 'Impractical' Generalizing Sociological Theories." *Sociology and Social Research.* 47:34-44.

Stouffer, Samuel, and Jackson Toby. 1951. "Role Conflict and Personality." *American Sociological Review.* 56:395-406.

Sumner, William Graham. [1906] 1960. *Folkways: A Study of the Sociological Importance of Usages, Manners, Customs, Mores, and Morals.* New American Library.

Vance, Rupert B. 1966. "Introduction to Jack E. Weller, *Yesterday's People: Life in Contemporary Appalachia.*" University of Kentucky Press.

Weller, Jack E. 1966. *Yesterday's People: Life in Contemporary Appalachia.* University of Kentucky Press.

Sex and Race Homogeneity in Naturally Occurring Groups*

BRUCE H. MAYHEW, *University of South Carolina*
J. MILLER MCPHERSON, *University of Arizona*
THOMAS ROTOLO, *Washington State University*
LYNN SMITH-LOVIN, *University of Arizona*

Abstract

We generate a number of hypotheses about face-to-face groups using the energy distribution principle: the frequency of an event is inversely related to the amount of energy expended in that event. The principle predicts that (1) the size of groups will be inversely related to the frequency of their occurrence; (2) at any group size, the composition of social positions will be less heterogeneous than chance; and, (3) as group size increases, observed compositional homogeneity will decline at a slower rate than chance. We test these hypotheses using data on more than 100,000 naturally occurring, public, face-to-face groups gathered in sampling sweeps through two communities over a three-year period. The data support the hypotheses and yield interesting differences in the strength of sex and race heterogeneity. We discuss the findings as they relate to the general energy distribution principle and to other sociological perspectives.

We begin with a general principle drawn from the field of thermodynamics called the *energy distribution principle.*[1] This principle is expected to hold in any system where the total amount of energy does not change markedly during a

** Bruce H. Mayhew died on 3 March, 1988, after a two-year battle with cancer. He had collected all the data analyzed in this article and written most of the theoretical argument at the time of his death. The remaining three authors, representing two successive generations influenced by him, have attempted to analyze the data, report the results, and draw conclusions from them in a manner consistent with the quality and character of Mayhew's thought. (It is typical of his confidence and foresight that he had already drafted parts of the conclusions section, assuming that his theoretical ideas would be supported by the forthcoming data analyses; therefore, even these parts of the article bear his imprint.) We intend this article as a tribute to his remarkable intelligence and as an additional contribution of his unique perspective on sociology. Since Bruce did not have the opportunity to complete his thinking about the project and to view the results of the data analyses before completing this article, the remaining three authors are, of course, responsible for any shortcomings of the current version.* *Direct correspondence to J. Miller McPherson, Department of Sociology, University of Arizona, Tucson, AZ 85721. E-mail: mcphersn@aruba.ccit.arizona.edu.*

 Social Forces, September 1995, 74(1):15-52

specified time interval, e.g., one day, one week, one year, etc. Given this condition, or a reasonable approximation thereto, the principle states that:

The frequency of an event is inversely related to the amount of energy expended in that event.

That is, if we observe a large number of events of type X, and if there is variation in the amount of energy expended in these events, then we hypothesize that a relatively large number of these events will involve relatively small energy expenditures and a relatively small number of these events will involve relatively large energy expenditures.

To take an example outside of sociology, consider the frequency distribution of earthquakes, classified according to the amount of energy in each earthquake as measured on the Richter scale. For the earth's crust within a time interval of one year, we find the general principle holds: large earthquakes occur infrequently, small earthquakes occur often.

Before turning directly to sociological substance, we need to be clear about relevant background conditions. We said that in order for the energy distribution principle to hold, the total amount of energy in the system should not change over the time interval studied. Or, at least, the amount of energy should be approximately constant over that interval. Is this true of events occurring in the earth's biosphere?[2]

To know whether there is significant variation in the amount of energy in the earth's biosphere in an interval of time chosen for study, we need to know whether the amount of energy emitted by the sun is variable or constant. That is, the vast majority of all energy in the biosphere comes from the sun. Although it was not known with any degree of certainty until as recently as 1973, variation in the amount of energy emitted by the sun is usually less than 1% (see Frazier 1982). And, since a variation of as much as 2% would precipitate an ecological catastrophe of biblical magnitude, we can be reasonably certain — from the fact that at least a few representatives of the human species are still alive — that no such event has occurred in recorded history. In any case, with the amount of energy pouring out from the sun varying less than 1% per year, it is reasonable to say on the basis of present evidence that the amount of energy in the earth's biosphere changes hardly at all over the course of a year. Accordingly, the energy distribution principle should apply to events in the earth's biosphere for an interval of one day, one week, one month, or one year.

Even if the total amount of energy entering the earth's biosphere does not change very much in one year, it is nevertheless true that the energy distribution over that biosphere is quite uneven (for details, see Schneider & Londer 1984:189-204). There are sudden, localized energy shifts in the form of earthquakes, volcanic eruptions, tornadoes, etc. These are clearly devastating to local populations, disrupting and even destroying whole communities. That these localized disasters are the exception, rather than the rule, permits social life to flow on and permits us to speak of the routine of everyday life. That routine would be impossible in the environment of frequent, sharp shifts in the energy level. Finally, the amount of energy flowing through human communities varies from one locality to another; large cities (through both their population and technology) harness larger amount of energy than do small towns.[3] What

matters for us is the relative infrequency of disasters and the relative stability of the energy level in each human community. If we are willing to assume that the routine we see is an indirect indication of highly stable energy levels, we can apply our energy distribution principle to most human communities, most of the time, and encourage students of disasters to fill in the gaps.[4] Without denying that disasters may have considerable relevance to the operation of human social systems, we shift to questions of more immediate sociological concern.

Our objective throughout will be to indicate how our energy distribution principle applies to each type of event discussed. First, we must establish clearly the scope of the principle's application. We are attracted to the principle because of its extreme generality: it describes the distribution of all events, both physical and social. We need to make clear, however, that it applies only to *events*. Sociologists often find themselves studying the characteristics of individuals rather than social forms (for more on this distinction see Mayhew 1980, 1981; Simmel 1894). People are not events, however. Individuals do not expend energy; their activities (social and nonsocial) do.[5] It is the distribution of activities, especially of social activities that create observable social forms, to which the sociologist will apply our principle.

As with any social form, measurement of events and their energy expenditure can be difficult. Part of the reason that sociologists so often study individuals rather than social phenomena is that they appear well-bounded, vivid, and easy to measure using conventional survey technology (Mayhew 1980, 1981; McPherson 1982). The boundaries of events can be somewhat difficult to determine (see our discussion in the Measurement section), but this problem is not unique to social phenomena. The boundaries of hurricanes and other lesser storms are ambiguous, but we can measure their energy expenditure by the millibars of pressure at their center (the lower the pressure, the more powerful the storm). Similarly, voluntary associations may have an amorphous membership that makes their size difficult to determine. However, we find that multiple measures of the size concept (such as an officer's report, a member's report, and an actual count of one meeting's attendance) all converge on a common value (McPherson & Rotolo 1995). While boundary and energy measurement issues may be complex for social phenomena, we argue that consistent definitions developed and applied to a particular domain of events are possible and are preferable to the alternative of restricting our study to individual characteristics. We now illustrate our thinking with the application of the energy distribution principle to a domain of social interaction.

The Size Distribution of Face-to-Face Groups

Consider the size distribution of face-to-face groups that may be observed in public places. Size refers to the number of people in the group. And, although groups are by definition never smaller than two persons, we will admit a more general distribution of *aggregates* that includes size one: a single individual appearing in public alone. If we are willing to assume that the number of people in an aggregate (group) is a rough measure of the amount of energy

being expended by the group, application of our principle is relatively easy in this case. That is, we would predict that aggregates of size one will occur most often, that aggregates of size two will occur less often, and that the largest groups will rarely be seen.[6]

Before passing beyond this case, let us note exactly how it is reasonable to equate the number of persons in an aggregate with increasing energy expenditures. If a single individual goes out into a public place — perhaps on the way to a movie or to go shopping — she or he expends energy walking, driving, or even just sitting or standing. Now, for aggregates of size two or larger, an additional consideration comes into play. As we choose to observe only those aggregates (of size two or larger) that are engaged in face-to-face interaction (rather than just being physically proximate), the energy expenditure for these larger aggregates will be compounded by the addition of each person not only because more individuals require more energy to operate as biological organisms, but also because they are engaged in social communication to maintain coordination of their actions (in shopping, attending a movie, going to lunch, etc.). This means that energy must be expended not only on going out in public (as would be true for any lone individual), but also on coordinating action to carry out the activity in which the group is engaged. Any cooperative effort requires that individuals transmit signals to one another, that each monitors the signals and behaviors of the others, and that each attempts to align his or her actions with those of the others in the group. Thus an excursion in a group increases each person's energy expenditure by virtue of the constant checking of signals among persons.[7] And, the larger the group becomes, the larger the number of others one has to monitor and coordinate action with. Hence, the energy expenditure would increase multiplicatively, rather than in simple arithmetic fashion.[8]

We do not argue, of course, that size (or its multiplicative function) is a perfect measure of energy expenditure. Two people of different physique and different rates of metabolism may not expend *exactly* the same amount of energy even though they are engaged in identical activities (e.g., walking, talking). Each dyad in a large face-to-face group may interact less intensively than the single dyad in a group of size two. In very large gatherings, a division of labor may occur that lessens energy expenditures.[9]

Before leaving the case of face-to-face groups and their energy expenditures, we wish to emphasize one point. We do not wish to imply that individuals *choose* to travel alone in public because it is more energy efficient. The energy distribution principle does not depend on a mechanism of human choice; if it did, it could not apply effectively to physical phenomena like earthquakes and storms. Rather, it is a general law or principle that describes the distribution of any domain of events that fit the conditions described in our initial statement.[10]

Social Interaction and Physical Distance[11]

Physical distance has generally been considered to regulate social contact because of the amount of energy that must be expended to overcome the friction of space. Longer distances require more energy, shorter distances

require less energy. Hence the distribution of interaction events is expected to follow our energy distribution principle: interactions over short distances (requiring small energy expenditures) will occur often, while interactions over long distances (requiring large energy expenditures) will occur much less often. This observation has been so well documented that Mayhew and Levinger (1976) have called it the "law of distance interaction," which states that "the likelihood of interaction or contact of any kind between two social elements is a multiplicatively decreasing function of the distance between them, or of the costs of overcoming that distance"[12] (93).

Note that in this argument, physical distance is used as a proxy for energy expenditure. As with group size, distance is likely to be an imperfect measure. When technology reduces the energy expenditure necessary for communication over longer distances, physical distance becomes a poorer measure (as when it takes less energy to call someone in Singapore than to contact someone next door who has no telephone service). If technology produces enough disturbance in the system that physical distance is no longer a reasonable proxy for energy expenditure, we might have to restrict the domain that could be studied with this proxy to face-to-face interaction. Notice that Mayhew and Levinger (1976) anticipated this qualification by explicitly including energy expenditures in their statement of the law of distance interaction. They note that changes in the "level of contact technology" will result in step-function effects in reducing the impact of distance on interaction (see also Hawley 1971). However, they argue that increases in technology have never reduced the effects of distance to zero in any society, and that therefore, the law of distance interaction should hold *ceteris paribus*. Here, we merely note that this law is a special case of our energy distribution principle and hence is subsumed under it.

Language and Social Interaction[13]

While not all contacts between humans involve information transfer in symbolic form (Mayhew 1984a), a great deal of what we call social interaction does involve such transfers through codes (natural languages) that have evolved over relatively long periods of time.[14] Furthermore, the amount of time and energy required to learn one of these codes is considerable. Hence, we expect that most people would learn only one or at most a very few such codes. Most people are monolingual. And, to the extent that communication requires a common code, language differences represent a barrier to communication, while language similarities provide low-energy opportunities for contact. It requires less energy to communicate with those who speak our own language, and much energy to communicate (or in any way coordinate action) with those who speak a different language. The more similar two languages are, the less energy required to overcome communication difficulties, and the less similar two languages are, the greater the amount of energy required to execute any kind of contact or coordination in interaction. So, the tendency of people to communicate most often with persons sharing the same language and less often with those who speak a different one is a special case of our energy distribution principle.

Social Interaction and Social Distance

One hypothesis that has developed in the sociological literature over many years might be called the social distance interaction hypothesis. It claims that

> The frequency with which people enter into face-to-face association (contact, interaction) is inversely related to the social distance between them.[15]

This homophily of social contact is one of the best-supported empirical generalizations in the sociological literature. Researchers have found homophily in all age groups (from schoolchildren to the elderly). It occurs on many dimensions, including basic sociodemographic characteristics such as age and sex; acquired characteristics such as education, prestige, and social class; personal attributes such as attitudes, beliefs, and social behavior (see a review of this voluminous literature in McPherson & Smith-Lovin 1987). Far from being a matter of individual choice, the increasing frequency of association between similar others often has structural sources (Blau 1977; McPherson & Smith-Lovin 1982, 1986, 1987).

In this article, we define social distance as any *difference* in social position. These include differences in age, sex, occupation, race, ethnicity, rank on wealth, rank on power, rank on prestige, years of education, and membership in any of a wide variety of social groups including families, organizations, etc.[16] All of these are social positions. If two individuals occupy the same position on a given dimension (e.g., sex), the social distance between them is zero on that dimension. If two individuals occupy different positions on the same dimension (e.g., wealth), then there is a nonzero social distance between them, although how we choose to measure that distance may vary according to a variety of theoretical rationales. At a minimum we can say that social distance can take on values of *zero* (when two people occupy the same position) and *one* (when two people occupy different positions). Hence, the social distance interaction hypothesis can always be tested at this minimal level of measurement.

If we identify a dimension called *sex*, with categories male and female, and another dimension called *race*, with categories African American and white, we can suggest a measurement extension for social distance, a measure that simply adds up the social differences in question. Thus for persons who interact with those like themselves, interaction is occurring across social distance zero. For persons who interact with others differing on only one dimension, interaction is occurring across social distance one, and for persons who interact with others differing from themselves on two dimensions, interaction is occurring across social distance two. Clearly, as long as we stay with two-category dimensions, the social distance scale can be indefinitely extended beyond these scores by adding new dimensions (e.g., age: old, young; wealth: rich, poor; power: strong, weak, etc.). It is important to realize that this illustration is a direct extension of the reasoning employed in our more elementary measure mentioned above. That is, if sex and race are social dimensions and their categories of male-female and African-American–white are different positions, we have in this extension merely operated on our original assertion that there is social distance between persons occupying different social positions. Again, with appropriate theoretical rationales, still different and even more refined measures can be suggested.[17]

Now that we have made clear (we hope) just what is meant by social distance, the social distance interaction hypothesis should be sufficiently clear to permit addressing the issue: Is this hypothesis a special case of the energy distribution principle? The answer is yes, and the reasons for this answer are outlined in the following paragraphs.

Humans are biological organisms that expend energy in overt and covert activity (covert activity refers to electrochemical energy expended in the central nervous system). Humans are allocated to different social categories (positions) that structure the experiences of their entire life courses, that is, that determine what they will learn and experience in social life. Their central nervous systems encode social information of various kinds, including behavioral programs associated with (1) their own social categories (male, white, young, attorney) and to a lesser degree (in some cases amounting to nearly zero) with (2) other social categories (opposite sex, different racial or ethnic groups, different age groups, different occupations, etc.). When humans form groups, the activity that occurs there requires the expenditure of overt and covert energy, especially in any case where joint action of members is required (to go on a shopping trip in a public place, etc.). Since humans have encoded more information about their own categories' expected behaviors, more time and energy must be expended in heterogeneous groups just checking signals for action (see Carley & Palmquist 1992). This time and energy expenditure diverts time and energy from the conduct of external joint operations (going shopping). So humans will in all sizes of groups (two and larger) form associations that are disproportionately homogeneous with respect to social position than would be expected by chance. People will associate most often with others like themselves because doing so decreases energy expenditures. People will associate least often with others different from themselves because doing so increases energy expenditures. If this is so, the social distance interaction hypothesis is a special case of the energy distribution principle.

The argument we have outlined for social distance may be amplified by recalling our discussion of natural languages. In effect, we have been saying that when people are socialized into their social positions, they learn a language (code) peculiar to those positions. The network ties that define their position lead to shared knowledge, a subcultural code, with others to whom they are tied (Carley 1991). While these languages are a part of the subculture of each position, rather than natural languages strictly defined, their effect is the same: they pose an *energy barrier* to communication. Just as it is difficult (requires great energy expenditure) to communicate or in any way coordinate action with others who speak a different natural language, so it is difficult (requires great energy expenditure) to communicate with persons who have learned the language of positional subcultures other than our own.[18] So, while speaking the same natural language (e.g., English), the subcultural language of an African American female attorney and the subcultural language of a white male dishwasher are sufficiently alien in the grammar of actions and expectations they permit that they constitute energy barriers to communication and coordinated action. We are saying that each position has its own code (language), that covert and overt actions are structured differently for different positions so that the social distances between them are virtually identical in form (if not in

strength) to the natural language barriers discussed earlier. Hence, the same energy argument applies, and, of course, the same deduction is appropriate: interaction regulated by social distance is a special case of the energy distribution principle. People will be observed in face-to-face associations with others like themselves quite often because less energy expenditure is required to sustain communication and coordinate action. People will be observed in face-to-face interaction with others unlike themselves only very rarely because more energy expenditure is required to sustain communication and coordinate action. At all sizes of face-to-face groups, there will be a bias toward positional homogeneity of the participants.

Heterogeneity in Face-to-Face Groups

We have already said that the size distribution (even including aggregates of size one) of face-to-face groups is a phenomenon we wish to understand. And we have claimed that at all sizes of face-to-face groups, there will be a bias toward homogeneity that may vary as a function of group size. That is, even if there is a general bias toward homogeneity at all sizes, this in no way excludes the possibility that homogeneity-heterogeneity may show systematic variation as a function of size.

If each increment in size places an added energy burden on communication between group members, as we have argued, then any increase in heterogeneity with size, or even the same level of heterogeneity at larger sizes, would compound the difficulties in communication and coordination of action. Therefore, we expect that the multiplicatively increasing relational coordination demands of larger groups will lead positional homogeneity to increase at an increasing rate with group size.

Hypotheses

Although we have discussed a variety of issues, only a few lend themselves to direct tests on groups observed in public places. We shall examine the following hypotheses:

Hypothesis 1. The size of groups (including aggregates of size one) will be inversely related to the frequency of their occurrence.

Hypothesis 2. In any size of group, the composition of social positions will be less heterogeneous than chance.

Hypothesis 3. Homogeneity will increase at an increasing rate with group size.

These hypotheses are made for face-to-face groups observed in public places. This domain satisfies the scope conditions of the energy distribution principle: it deals with *events* for which we can observe reasonable (if imperfect) measures of energy expenditure — the size distribution and heterogeneity of the groups.

Data Collection

"Each discipline tends to develop a distinctive way of characterizing the problems it confronts — its own paradigm or world view. Fields vary also with respect to the forms in which their typical data are observed, measured and recorded" (Freeman 1984:346). While Freeman's statement is certainly correct, there are also some similarities in the problems confronted by different fields of study. Marine biologists studying hammerhead sharks in oceans and sociologists studying face-to-face groups in public places share the difficulty of identifying the population from which a sample can be drawn. There is no known population of hammerhead sharks in the oceans and there is no known population of face-to-face groups in human communities. So both marine biologists and sociologists are forced to adopt a procedure of making "sweeps" through areas where they suspect they may find "representative" cases. The rationale for this procedure (aside from lack of choice) is that after a large number of studies have been conducted by different investigators in different locations, the total amount of data collected and analyzed in all the studies will approximate — rather than completely misrepresent — a reasonable sample from the universe of cases. On this optimistic note, we turn to our data collection procedures.

SITE CHARACTERISTICS

Two towns in South Carolina were selected for making observational sweeps. The first, Winnsboro, is a small town that serves as both a market and an administrative center for Fairfield County. Winnsboro is the county seat and occupies a central geographic location in the county. Fairfield County is extremely rural, with about 80% of its land area in forests and fields. Winnsboro is the largest population concentration in this county; in 1980 Winnsboro's town limits contained 2,919 people, although local government authorities estimate that Winnsboro and its immediate environs has a population of around 8,800 people (Green 1985). The entire population of Fairfield County was 20,700 people in 1980. The second town, Columbia, is the state's largest city, the state capital, and the county seat of Richland County. Columbia is the central city of a two-county SMSA consisting of Richland and Lexington counties. Our observations for Columbia were confined to its urbanized area within Richland County. Within its city limits, Columbia had a population of 101,208 people in 1980, while Richland County had a total of 269,735. The entire metropolitan area (Richland and Lexington counties) had a total population of 410,088 in 1980.

There are two principal highways linking Winnsboro and Columbia. The first separates them by 29 miles, the second by 34 miles. The northern part of Richland County, which forms its border with Fairfield County, resembles Fairfield County generally in being composed primarily of forests and fields. Consequently, there is a rural region approximately 20-25 miles wide separating Columbia and Winnsboro. The two towns are quite distinct populations separated by a wide rural space.

OBSERVATIONS

In each town, observations were conducted between October 1979 and August 1982. The observation periods, or "sweeps," were carried out for each morning and afternoon for each day of the week in both towns. This sequence was repeated four times, producing a total of 56 data sets for each town ($2 \times 7 \times 4 = 56$) for a total of 112 observation periods. Most sweeps represent a two-hour observation period. The number of observational sweeps were balanced by month of the year, as well as day of the week. In Winnsboro the observations were made by sweeping through the entire town and also in one shopping center located approximately one-half mile from the town limits. In Columbia the vast majority of the observations were made in a 52-block area comprising the principal business district, the state capitol and office buildings, and a large university; at the state fair grounds (located within the city limits), and at two large shopping malls located outside the city limits but within the urbanized area.

During a typical observation period an observer walked or drove from one location to another. The observer stopped periodically at street corners and made an observational sweep noting the number of groups (including one-person aggregates) and the sexual and racial composition of each aggregate. The observer then moved on to another location. These steps were repeated over a period of two hours. In making observational sweeps, the observer made an effort to distinguish actual face-to-face groups from persons who just happened to be standing close to one another. In a large number of cases (about half), this presented no problem because many groups arrived on the scene together in a car.[19] In other instances, the observer made an effort to ensure that the people were actually talking together before including them in the observation set. Also, to minimize errors, the observer confined observations to those collections of people within 20 meters of the observer.[20]

Results

Two outcomes of our observations in Columbia and Winnsboro require emphasis. First, for every observation period all categories of persons (white male, white female, African American male, and African American female) were present in the observation area and appear in each data set. Second, every type of dyadic (or pairwise) interaction within and across social categories was observed at one time or another in both towns. Thus, Blau's (1977) two axioms, that people interact with others like themselves and that people interact with others different from themselves, are empirically correct for the two towns.

METHODS AND RESULTS FOR HYPOTHESIS 1

Hypothesis 1 predicts that the occurrence of groups will be inversely related to size. The frequency distributions for the size of the naturally occurring groups for Columbia and Winnsboro are shown in Table 1.

As one can easily see, the peak of the distribution for both the towns occurs at size one. As size increases, the distribution shows a sharp decline. This

TABLE 1: Frequency Distribution of Group Size by Town[a]

Size	Winnsboro Frequency	Columbia Frequency
1	24,839	43,799
2	10,630	16,347
3	3,375	3,854
4	1,237	1,463
5	514	449
6	256	218
7	56	59
8	33	28
9	11	18
10	10	6
11	3	3

[a] Data gathered by observation in a total of 56 random sweeps of each town, balanced by month of the year and day of the week. See further information in text.

decline is emphasized by observing that more than 94% of the groups observed in Columbia and Winnsboro were of size three or less. Hence, as this table illustrates, the size distribution of aggregates in both towns conforms to the prediction of hypothesis 1.

Because our test of hypothesis 1 is based on the analysis of human groups, which are subject to constraints of coordination among the members of the group, we can make more specific predictions than simply that the observed frequency is a declining function of group size. Recall that the basic argument for this hypothesis is that the maintenance of the increasing number of dyadic relationships embedded in a group increases multiplicatively, rather than additively, as size increases. As we have discussed elsewhere (Mayhew & Levinger 1976:93-94), the potential number of contacts in an aggregate depends upon both the maximum possible number of contacts (the number of possible dyadic relations) and the underlying contact distribution (how many contacts each point or person has). In very large aggregates (e.g., cities or societies), we must make assumptions about the contact distribution in order to estimate the density of interaction — what Durkheim (1893:283) called "dynamic density" (see the discussion in Mayhew & Levinger 1976:94). But for small face-to-face groups, it is reasonable to assume that contact is often saturated and that most of the maximum potential number of contacts (or relations) are activated. To the extent that this assumption is not met, using the maximum potential number of contacts as a baseline model for the number of contacts in the group and our more specific version of hypothesis 1 developed below will overestimate the increasing rate at which size will affect energy expenditure. Since this question is an empirical one, we proceed with our development of a formal baseline with which to compare our frequency observations.[21]

Formally, the number of dyadic relations in a group of size S is:

$$D_s = \binom{S}{2} = \frac{S(S-1)}{2} = \frac{1}{2}S^2 - \frac{1}{2}S \quad (1)$$

where S is the size of the group and D_s is the number of embedded dyads. Values of D_s and S are presented, along with a graph of the structure of dyadic relationships, in Figure 1. As the figure shows, D_s is an increasing quadratic function of S. As the number of people in the group increases, the number of dyadic relationships in the group increases nearly as half the square of group size. According to the energy distribution principle, a given group will be constrained by the coordination effort captured by equation 1, as well as by the necessary energy expenditures for maintenance and transport of each individual in the group (see the discussion in an earlier section). We don't expect that the relationship in equation 1 will be mirrored exactly in the data, since probably not all possible dyads in the group will be activated for larger groups in a given time frame. What we expect is that the observed curve will be a member of the family of curves that increase quadratically; the curve should accelerate. Thus, we do not argue that the distribution of group size will exactly mirror the function described in equation 1 (that is, the coefficients will not exactly equal ½), but it should have the same general form. We can write equation 1 in the form of a regression model of counts:

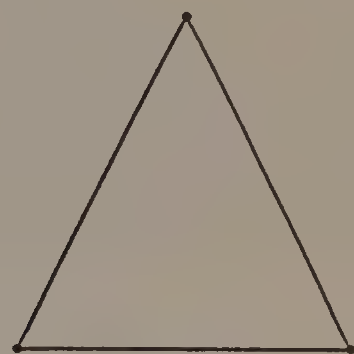
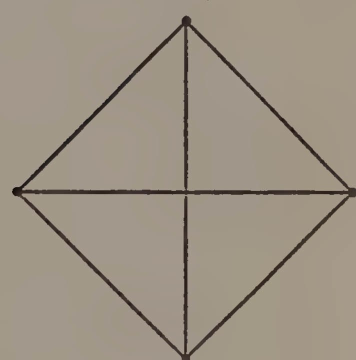
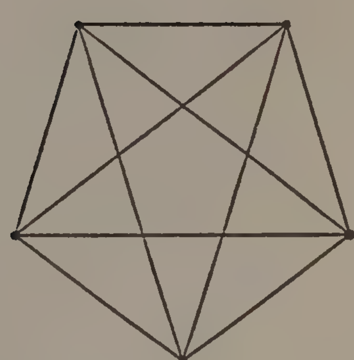
$$N_s = \alpha + \beta_1(Size) + \beta_2(Size^2) + \gamma \quad (2)$$

where N_s is the number of groups of size S, and α, β_1, and β_2 are parameters to be estimated from our data. The qualitative prediction for equation 2 we derive from hypothesis 1 as interpreted by equation 1 is $\beta_1 < 0$ and $\beta_2 > 0$. These predictions follow from the signs of the coefficients of S and S^2 in equation 1. When $\beta_1 < 0$ and $\beta_2 > 0$, size and frequency are related multiplicatively according to the general form of equation 1. We estimate the parameters in equation 2 with negative binomial regression analysis.[22] As Table 2 shows, this hypothesis is generally supported. Although the coefficient for $size^2$ is not significant in Winnsboro, all coefficients are in the predicted direction. The results strongly argue that the general relationship between group size and frequency of occurrence is of the multiplicative form predicted by our elaboration of hypothesis 1. Not only is frequency of occurrence a strongly decreasing function of group size ($\beta_1 < 0$), but the exact shape of the curve relating size and frequency is consistent with the idea that the number of embedded dyads in the group adds to the energy expenditure necessary to coordinate the group ($\beta_1 < 0$ and $\beta_2 > 0$).

METHODS AND RESULTS FOR HYPOTHESIS 2

Hypothesis 2 states that conditioning on group size, the composition of social positions will be less heterogeneous (i.e., more homogeneous) than chance for any group of at least size 2. A test of this hypothesis requires that we calculate the expected number of homogeneous groups under random expectation for the two types of social composition in both towns. In other words, we require a baseline to compare with our observed results (Mayhew 1984a). This baseline will show what random associations of individuals would produce and allow us

FIGURE 1: The Number of Dyads Embedded in Groups of Increasing Size

Size (S)	D_s (Number of embedded dyads)	Structure of dyadic relations
2	1	
3	3	
4	6	
5	10	
N	N(N-1)/2	—

to ascertain what systematic sociological forces (such as the energy distribution principle) might be operating to produce deviations from a random assortment of individuals.

To obtain the baseline expectation, we use the definition of the expected value of a random variable. We calculate the probability of a homogeneous group for a group at each size, where size varies from 2 to 11 (the observed maximum size in our data). Since the social composition variables sex and race are both binary response variables, the binomial distribution governs the probability of a homogeneous group of size S forming under random expectation.

In order to understand why the binomial distribution is the correct baseline, it is helpful to recall how the binomial distribution can be used to calculate expected values for the familiar experiment of tossing a coin n times (where $n = 1, \ldots N$). Consider the experiment of flipping a coin three times. We know p = Pr(Head) = 0.5 and q = Pr(Tail) = $(1 - p)$ = 0.5. Hence in order to calculate the probability of either exactly three heads (e.g., corresponding to a homogeneous group of three males) or exactly three tails (e.g., a group of three females, we compute

$$Pr(3\ Heads\ or\ 3\ Tails) = (0.5)^3 + (0.5)^3 = 0.25$$

The situation for our research problem is similar to that of the coin-tossing experiment. In our case, the size of the group under the random baseline is analogous to the number of coin flips. However, we cannot simply assume that the probabilities of the binary response for either race or sex are one-half, since the total number of people with a given social composition differs for each town. For example, the towns are not composed of a population of exactly 50% male and 50% female. Instead we use the observed proportions for each social composition variable. For instance, for sex homogeneity, we determine the proportion p of females in each city by pooling all our group observations for that city.[23] Then we can define $q = (1 - p)$, where q is the proportion of males in each city. These proportions represent the probability of encountering a female or male in the public places of each town. Similar proportions can be calculated for race homogeneity in each town. From this information and the frequency values in Table 1, we can determine the probability of observing homogeneous groups under random expectation at each group size. The expected number of homogenous groups at each size is

$$E(Homogenous\ group_i) = n_i Pr(Homogenous\ group_i) \qquad (3)$$

where n_i = total number of groups observed at group size i ($i = 2, \ldots 11$).

Hypothesis 2 predicts that the composition of social positions will be more homogeneous than chance. The results for hypothesis 2 are shown in Table 3.

As these tables show, for any group, the observed number of homogeneous groups is larger than the expected number under the baseline model. Groups in both communities are more homogeneous with respect to sex and race than would be expected by chance.

To illustrate this fact concretely, let us examine the racial composition of large groups in Winnsboro (shown in the first panel of Table 3). Given the roughly 50-50 representation of African Americans and whites in the communi-

TABLE 2: Test of Hypothesis 1: Parameter Estimates[a] for Size and Size^2 for Negative Binomial Regression Model[b]

	Dependent Variable Frequency of Group Occurrence	
	Winnsboro	Columbia
Constant	11.1450	11.9655
	(0.0767)	(0.0930)
Size	-1.0123	-1.2862
	(0.0704)	(0.0856)
Size^2	0.0104	0.0362
	(0.0122)	(0.0145)
γ	3.0387	4.0584
	(0.5989)	(0.5467)
Log likelihood	350192.82	607220.37
Degrees of freedom	11	11

[a] γ is a stochastic component that accounts for possible overdispersion in event count data.
[b] Parameter estimates for size and size^2 were calculated using the negative binomial regression method in COUNT (King 1989a). Standard errors for the estimates are in parentheses.

ty (see note 24), it would be extremely unlikely that a large group would not include at least one person of each race if the energy distribution principle were not in effect. Indeed, the baseline model predicts that we would be unlikely to see even one group larger than size 7 that was homogeneous on race; the expected value for a group of size 8 or larger is .37 (adding .30, .05, .02, and .00). Such groups should be extremely rare by chance. But our data show that 53 such racially homogeneous groups are observed, where none (or at most one) would have been expected. Therefore, we believe some sociological principle (such as the energy distribution principle) to be in operation.

METHODS AND RESULTS FOR HYPOTHESIS 3

Our test of hypothesis 3 uses the same baseline model as in hypothesis 2 in order to test the proposition that compositional homogeneity will decline at a slower rate than baseline expectations. We use the same binomial distribution to plot the probability of group homogeneity as a function of size and then determine the rate at which the function declines. Similarly, we model the observed probability of compositional homogeneity as a function of size and determine its rate of decline.

TABLE 3: Number of Homogeneous Groups for Winnsboro and Columbia: Observed vs. Expected

	Race Composition				Sex Composition			
	Winnsboro		Columbia		Winnsboro		Columbia	
Size	Observed[a]	Expected[b]	Observed	Expected	Observed[c]	Expected[d]	Observed	Expected
2	10,366	5,347.95	15,890	9,263.27	6,907	5,334.76	9,732	8,230.71
3	3,276	859.61	3,694	1,359.69	1,962	857.59	2233	983.54
4	1,189	160.44	1,392	336.64	497	159.20	584	190.48
5	489	34.13	435	69.15	182	33.92	182	30.04
6	249	8.76	201	22.76	81	8.65	70	7.52
7	54	0.99	51	4.19	23	0.97	15	1.06
8	32	0.30	24	1.36	6	0.30	14	0.26
9	11	0.05	15	0.60	2	0.05	5	0.09
10	8	0.02	5	0.14	1	0.02	1	0.02
11	2	0.00	1	0.05	0	0.00	3	0.00

[a] Sum of number of racially homogeneous groups (either all white or all black) at each group size.
[b] Expected number of groups under binomial baseline from equation 3.
[c] Sum of number of sexually homogeneous groups (either all male or all female) at each group size.
[d] Expected number of groups under binomial baseline from equation 3.

Since compositional homogeneity is a binary response (homogeneous group = 0; heterogeneous group = 1), we use a logit model of proportions (the continuous analogue to the discrete binomial in Table 2) to estimate the rate of decline of homogeneity with size. The model for the decline in homogeneity is the logit model (Hosmer & Lemeshow 1989):

$$\frac{1}{\left(1 + e^{-(\alpha + \beta * SIZE)}\right)} \tag{4}$$

Parameter estimates for the rates of decline for each of the four models are shown in Table 4. The estimates can be interpreted as the log odds of encountering a homogeneous group as compared to viewing a heterogeneous group in the data. In all models, the parameter estimate for size is negative and highly significant, indicating that as size increases the odds of observing a homogenous group decrease as compared to a heterogeneous group.

Figures 2-5 show graphical representations of the rates of decline in the probability of compositional homogeneity by social composition in both towns.

All four figures show conclusively that the observed probability of a homogeneous group declines more slowly than the expected rate under the baseline model, strongly supporting hypothesis 3. The results are significant beyond the 0.001 level. The formal tests of the differences, based on a nonlinear estimation routine in SAS PROC NLIN (SAS/STAT User's Guide 1989), are

TABLE 4: Parameter Estimates from Logistic Regression Models for the Relation between Observed Compositional Homogeneity and Size[a]

	Race		Sex	
	Columbia	Winnsboro	Columbia	Winnsboro
Intercept	4.043***	3.962***	0.935***	1.395***
Size	-0.259***	-0.154***	-0.265***	-0.386***
	(0.0321)	(0.0407)	(0.0161)	(0.0174)
-2 Log likelihood	53.954	12.823	283.438	535.842

[a] Models estimated with PROC LOGISTIC routine (SAS/STAT User's Guide 1989). The negative coefficients indicate that the log odds of observing a homogeneous group decline as a function of size.

*** $p = .0001$

presented in Table 5. These tests compare the rate of decline in observed compositional homogeneity to the rate predicted by the baseline model.

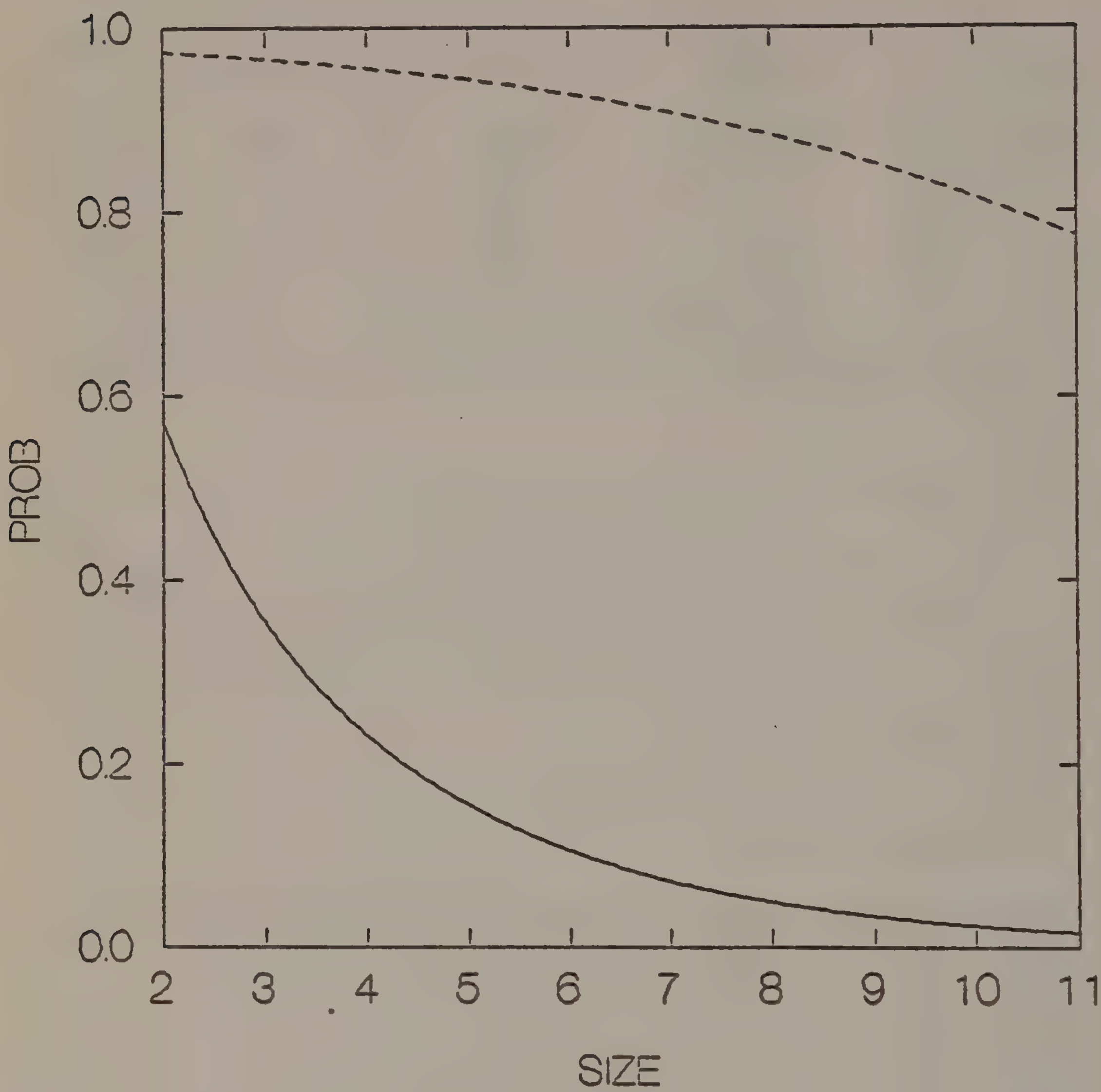
In all four cases, the *t*-statistic is extremely large, indicating that there is a significant difference between the observed rate of decline and the rate of decline under random expectation. The observed probability of a homogenous group declines much more slowly than the model of random expectation predicts, strongly supporting hypothesis 3.

A SUPPLEMENTARY FINDING

As may been seen in Tables 4 and 5 and Figures 2-5, the bias toward homogeneity at all sizes of groups is more pronounced for race than for sex. This is especially true for Winnsboro, where the difference between expected and observed homogeneity decline is the largest that we observe (see column 3 of Table 5). We did not hypothesize this outcome in advance, but we interpret this result as being consistent with the energy distribution principle on at least two grounds. First, this result might mean that the subcultural codes of racial groups are more different than are the subcultural codes of the sexes. If so, it means that there are, by our reasoning, greater energy barriers to communication between African Americans and whites than between males and females on language (subcultural code) grounds alone.

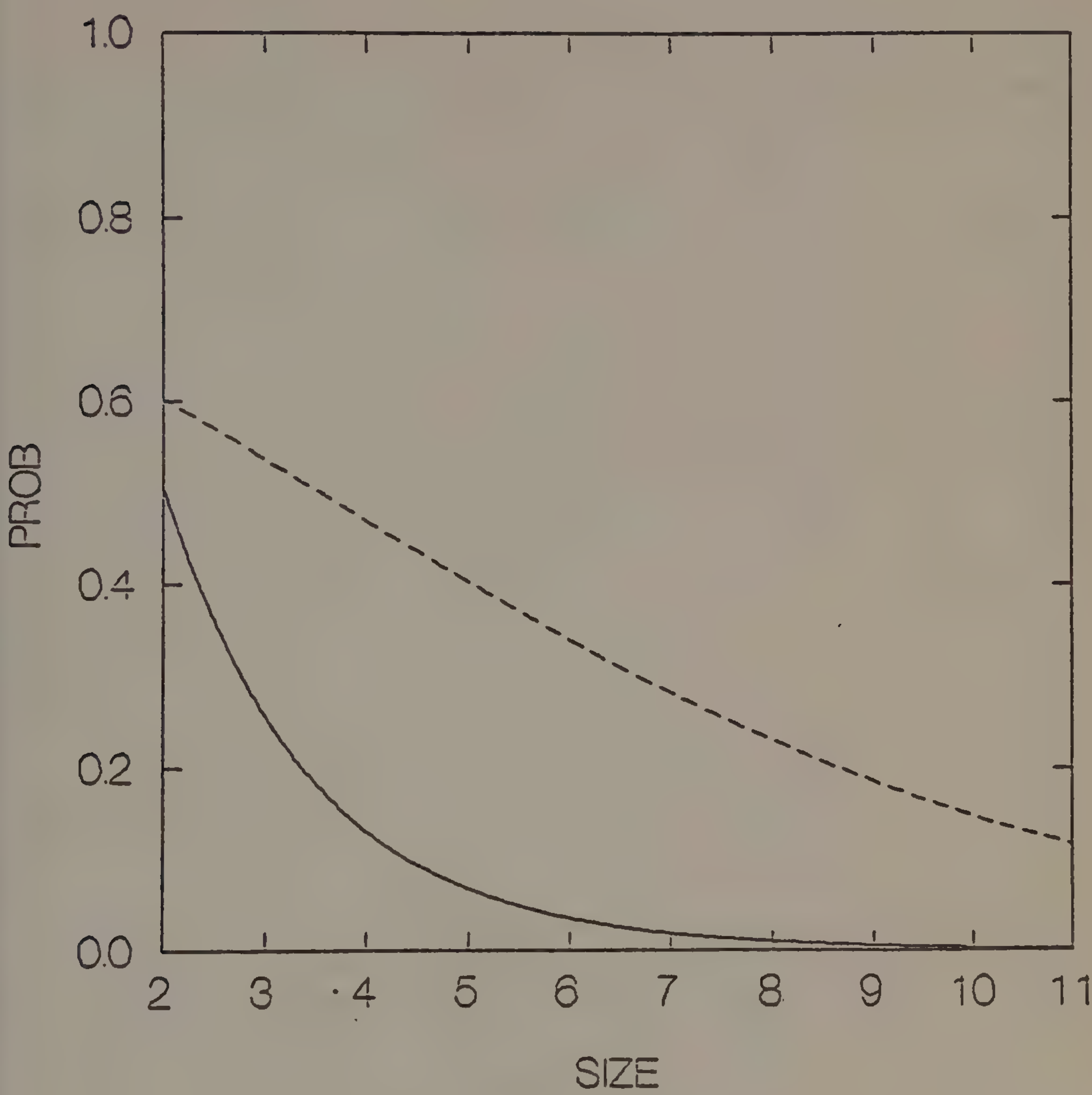
We might expect such a situation if race is more correlated with other dimensions than sex. Blau (1977), for example, has pointed out that nominal categories such as male and female, or African American and white, may derive their primary social significance from the fact that they are highly correlated with other social dimensions. In particular, they may be highly correlated with material wealth or income. Blau suggested that a great deal of the difference in social position for African Americans and whites may derive from their

FIGURE 2: Probability of Observing a Homogeneous Group as a Function of Size for Columbia Race[a]



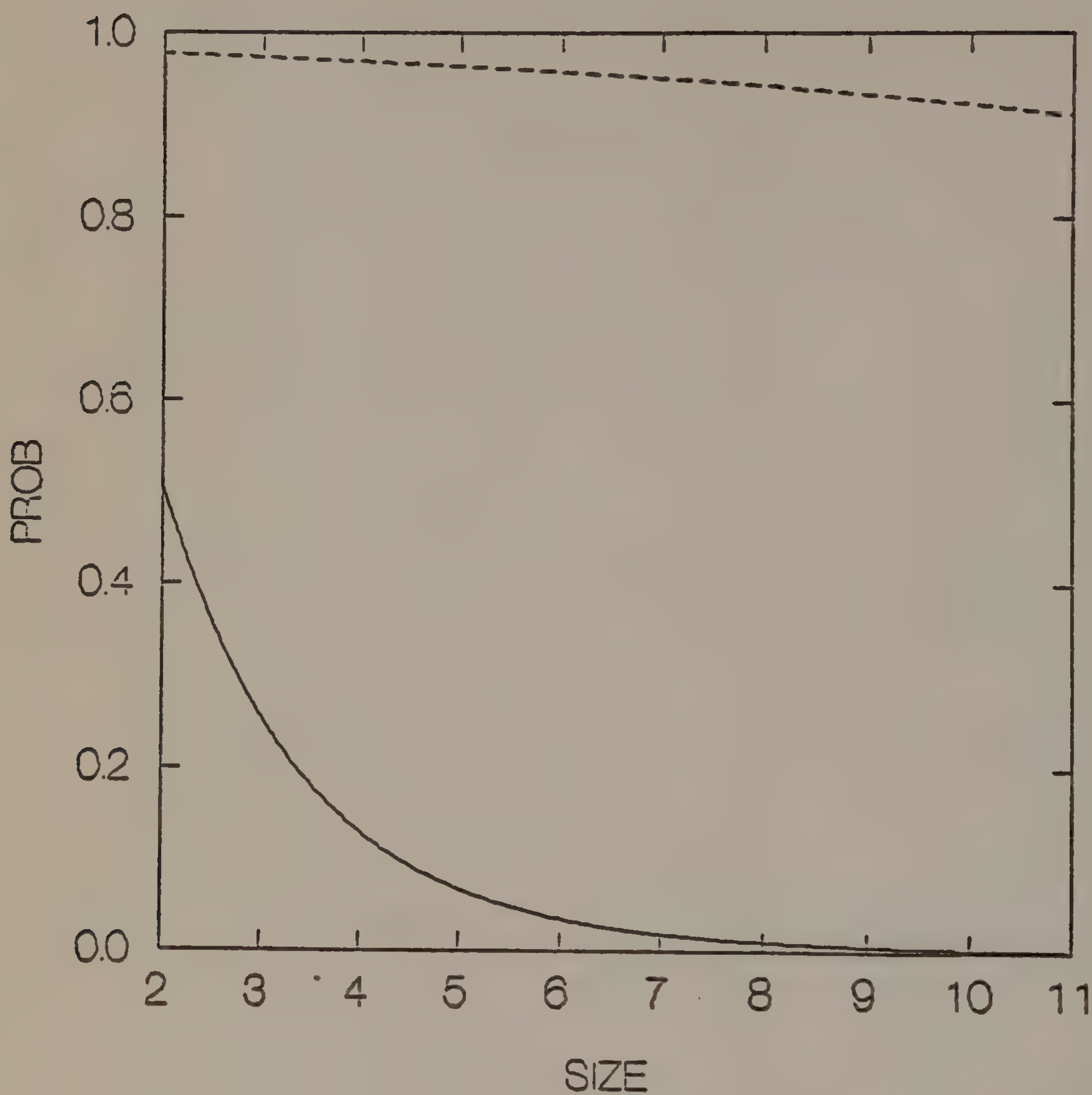
[a] Dashed line is the observed rate of decline for the probability of observing a homogeneous group. Solid line is the expected rate of decline for the probability of observing a homogeneous group.

FIGURE 3: Probability of Observing a Homogeneous Group as a Function of Size for Columbia Sex[a]



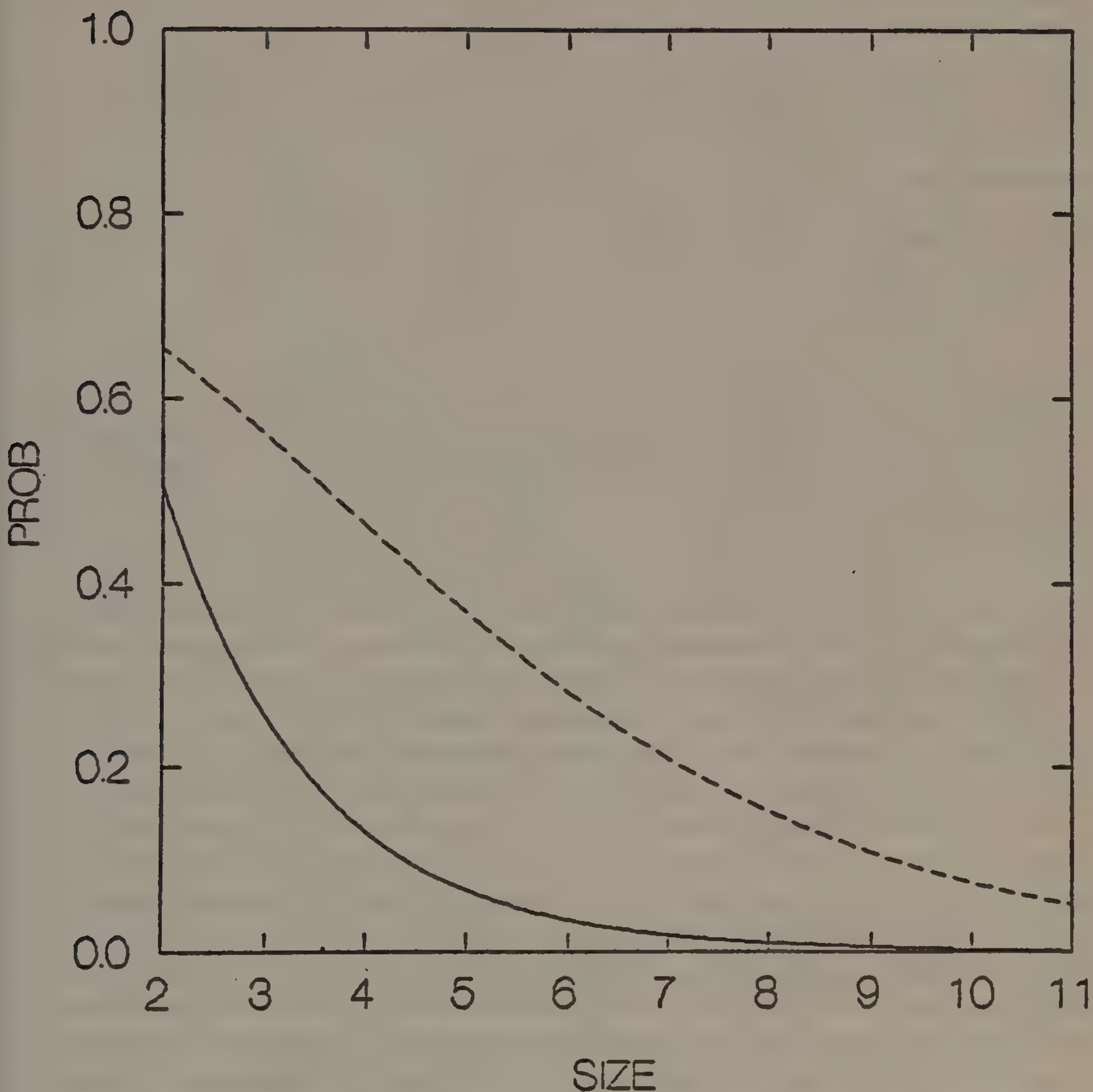
[a] Dashed line is the observed rate of decline for the probability of observing a homogeneous group. Solid line is the expected rate of decline for the probability of observing a homogeneous group.

FIGURE 4: Probability of Observing a Homogeneous Group as a Function of Size for Winnsboro Race[a]



[a] Dashed line is the observed rate of decline for the probability of observing a homogeneous group. Solid line is the expected rate of decline for the probability of observing a homogeneous group.

FIGURE 5: Probability of Observing a Homogeneous Group as a Function of Size for Winnsboro Sex[a]



[a] Dashed line is the observed rate of decline for the probability of observing a homogeneous group. Solid line is the expected rate of decline for the probability of observing a homogeneous group.

TABLE 5: Test of Hypothesis 3: Comparison of the Rate of Decline in Observed Compositional Homogeneity vs. the Expected Rate of Decline in Compositional Homogeneity under Baseline Model

	b[a]	β	b - β	t	P
Winnsboro rates					
Sex homogeneity	-.386	-.946	.560	32.18	< .001
Race homogeneity	-.154	-.944	.790	19.41	< .001
Columbia rates					
Sex homogeneity	-.265	-.941	.676	41.99	< .001
Race homogeneity	-.259	-.611	.352	10.97	< .001

[a] The observed rate of decline (b) was estimated with SAS Proc Logistic and the expected rate of decline (β) was estimated with SAS Proc NLIN. In all cases the t-statistic indicates that the observed rate of decline is significantly less that the expected rate of decline, supporting hypothesis 3.

differential rank on the dimension of income.[24] In the U.S., whites have higher rank (status) on median income than African Americans, and men earn more income than women. If this income difference is an extremely powerful factor in determining the life chances of persons in different positions — as Blau clearly believed — then such differences as we may observe for sex and race could easily be a function of their (median) income position and the life experiences that it shapes. These effects of income on life experiences (and therefore on subcultural codes) will be stronger for race than sex, if we consider that income is often translated into consumption and other experiences in household units. Men and women in the same household may share these experiences to some degree;[25] households, of course, tend to be quite homogeneous with regard to race. If income can have an impact on social interaction, or is in any way highly correlated with other position dimensions, greater observed effects for race as compared with sex may be operating primarily through rank (status) on income (material resources). This interpretation would allow us to generate hypotheses about regional differences that we would expect if our study in replicated in nonsouthern regions. To the extent that the American South has a higher correlation between race and other social positions, we would predict that the relative effects of race and sex would be more equal in other regions. In areas with more middle-class African Americans who are more widely spread through occupational, religious and voluntary systems, we would expect to find the subcultural codes of African Americans and whites to be more similar. The sex-heterogeneity of households probably will lead race to be more important than sex in most contexts; however, the relative size of the distance should decrease to the extent that whites and

African Americans occupy similar social positions. Informal observations in Arizona support this prediction.

A second factor also may be important in the relative weighting of race and sex in our analyses. It will be noted that, whereas the sexes are sprinkled relatively uniformly across virtually all residential areas, African Americans and whites are not.[26] The higher degree of residential segregation on race means that African Americans and whites are physically separated by greater distance that are the sexes. Hence, the law of distance interaction (a special case of the energy distribution principle) predicts that group homogeneity on race will be greater than group homogeneity on sex. This second argument presupposes that many of the groups we see go out into public have begun their journey at home and so may be composed of members of the same household or immediate neighbors. Both arguments may be correct: language and subcultural code differences between people are commonly a function of the physical distance separating them (Mayhew 1984b).

Notice that our argument about residential segregation by race would elaborate and qualify our discussion of regional differences above. Since the South (especially in smaller towns and cities) generally has somewhat lower levels of racial residential segregation than other regions, the law of distance interaction would lead us to expect a relatively *smaller* effect of race in the South than in other regions. A more complete analysis would take into account *both* the level of positional racial segregation (e.g., in occupations and income) and the level of residential segregation. The dramatic differences between the expected and observed homogeneity in Figures 2 and 4 may be an indication that the subcultural variation created by positional segregation may be a more potent force than the physical variation created by residential segregation. A systematic study that included variation in these factors and that observed groups in neighborhoods as well as central locations would be necessary to test this hypothesis.

Since such a study is not yet available, we continue our discussion of the logical implications of the hypotheses. If our arguments are correct, we also have a (post hoc) interpretation for the city differences that appear in Tables 4 and 5. Because of its rural, heavily agricultural character, Winnsboro and surrounding Fairfield County have a much greater correlation between race and other positional characteristics than exists in Columbia and its SMSA. In Fairfield County, most African Americans live in rural communities, working in either the logging industry or low-paying service jobs, or supported by social welfare programs. Whites are more likely to live in town, work at local small manufacturing plants, or occupy professional service jobs. While the differences between the races in Columbia also are dramatic, they are muted by the presence of state government, a major university, and the military; all these institutions create in-migration from outside the South and tend to be more egalitarian in their hiring than more traditional, local organizations. Given these factors, we believe that the subcultural and physical distances between the races are greater in Winnsboro than in Columbia, leading to our observation that racial homogeneity is more pronounced (as a declining function of size) than is sex homogeneity.

Some Implications

The energy distribution principle has served as the basis for deriving a number of predictions about face-to-face social contacts among humans. Only a few of these were translated into predictions about associations in face-to-face groups in public places, but the hypotheses so obtained were all confirmed in the research reported here. Now that we have at least one such confirmation, it is appropriate to summarize the general argument we have made. The law of distance interaction, the tendency of people to interact primarily within their own language group, and the social distance interaction hypothesis are all subsumed under the energy distribution principle. These, in turn, produce two of our hypotheses; the first hypothesis derives directly from the energy distribution principle. Finally, our finding that homogeneity is greater on the dimension of race than on the dimension of sex follows from the law of distance interaction.

ALTERNATIVE EXPLANATIONS FOR FREQUENCY AND HOMOGENEITY

For each of the three hypotheses that we have discussed, earlier researchers have proposed alternative interpretations. However, proposals that offer a different explanation for each hypothesis would be inferior to any proposition (such as the energy distribution principle) that subsumes all three interpretations. Furthermore, some alternatives are more apparent than real: there are many propositions that can be translated into energy distribution terms. We now turn briefly to some of those alternative explanations and show how they relate to our arguments.

Several models have been constructed to predict the *shape* of the group (aggregate) size distribution (cf. Cohen 1971; Coleman 1962; Coleman & James 1961; Goodman 1964; White 1962). For example, Coleman and James (1961; Coleman 1962) developed a mathematical model that argues that the rate of people joining and leaving a group is determined by the density of contacts within a closed system (i.e., a multiplicative function of the number of isolates and the number of groups). White (1962) used assumptions from queuing theory to develop alternative models with varying assumptions about the probability of joining and leaving groups. He dealt with both open and closed systems, and developed a model (the gamma model) in which the rate of joining and leaving depended on the number rather than density of isolates. All these models are based on the opportunity structure of contacts within the system and they predict the general functional form of the size distribution that we observe.

All these models are consistent with the energy distribution principle. They do not, therefore, constitute alternatives to hypothesis 1. Since they are structural explanations, based on the opportunities afforded by the distribution of contacts within the system, they are analogous to the physical theories that explain how chance configurations of weather patterns come together to form storms of various intensities. The important point is that the series of events (in our case, joining and leaving events) that will result in a large group will be rarer than the series of events that will lead a person to stay an isolate or to join

a smaller group. We also note that while these explanations provide an interpretation for the size distribution (hypothesis 1), they are silent on the homogeneity predictions of hypothesis 2 and hypothesis 3. The energy distribution principle is therefore to be preferred on the basis of both generality and parsimony.

Another class of theories that might speak to the size of human groups focuses on economies of scale and the supposed efficiencies that come from coordinating efforts among differentiated components (Smith 1791). A naive (and easily falsified) prediction from this division of labor — efficiency argument would be that large groups should be more common than individuals and small groups. A more sophisticated version of the argument would tie the size distribution to the goal of the group and the tasks being performed in the groups. Even if we assume that appearing in public places can subsume a wide variety of goals, this economic framework would have to posit a complex distribution of goals and tasks to explain the observed distributions of group size. In particular, the economy-of-scale argument would not predict the specific functional forms of our hypothesis 2 and hypothesis 3 predictions: why would tasks or goals be so arrayed as to produce economies in exactly this form? Since the energy distribution principle can predict the observed distributions with a single, extremely general law, it is preferable on those grounds.

Issues of homogeneity and heterogeneity (our hypothesis 2 and hypothesis 3) are often addressed by affective or cognitive theories in social psychology. These theories posit that people are more likely to be attracted to those who are similar to them on virtually any characteristic, including sociodemographic characteristics, achieved statuses, attitudes, or beliefs (e.g., see McPherson & Smith-Lovin 1987:370).

Cognitive theorists also have dealt with similarity issues. People are more likely to choose to associate with similar others as the result of social comparison processes (Festinger 1954). If people in public places were motivated to assess their own behavior for its appropriateness, these comparison processes could explain the homogeneity in our groups. If we are willing to make the further assumption that behavior appropriateness becomes a more salient issue if the group is larger, then this approach might even explain the tendency of our groups to get more homogeneous (relative to chance expectations) as they become larger. Modern cognitive theorists argue that the cognitive complexity of the knowledge associated with a domain influences the latency (processing time) for responding to stimuli in that domain (Eysenck 1984). If we assume that stereotypes of racial groups are not as clear in individuals' minds as gender stereotypes, or if individuals tended to have more counterexamples to racial categorization than to gender categorization, this greater cognitive complexity could explain our result that race homogeneity differs more from our baseline model than does sex homogeneity. These arguments obviously lack the power and generality of the energy distribution principle but might be interpreted as special cases in the following sense.

The affective theories that explain attraction can be translated into cognitive theories that focus on information load. People "like" things that are familiar, that require little processing load, and that allow a predictable environment. They "prefer" others whose positions and characteristics are similar enough to

make a reasonable basis for self-comparison and self-evaluation. All this may be summarized efficiently by stating that people like those who may be expected to have similar knowledge bases and cognitive structures (Carley & Palmquist 1992). Further, we believe that the cognitive arguments are consistent with and can be subsumed under the energy distribution principle. As we elaborated above in our discussion of natural languages and subcultural differences, more energy is required to decode and encode interactions with those who are different from us. The latency measure used by cognitive psychologists is explicitly interpreted as such: the longer the latency, the more cognitive network links must be traversed and the more processing energy used to decode a stimulus and encode a response to it. Therefore, the cognitive processing (and the affective affiliations that it generates) are special cases of our principle. Furthermore, we note that the affective and cognitive theories do not produce an explanation for the frequency distribution in hypothesis 1 (although the cognitive processing arguments might predict that the processing load would go up with the number of coordinating dyads — a prediction that elaborates our principle).

Explicitly structural theories that predict homogeneity in groups are Blau's work on inequality and heterogeneity (Blau 1977; Blau & Schwartz 1984), Skvoretz' (1983) elaboration of that work, and McPherson's (1983, 1990; McPherson & Smith-Lovin 1986, 1987) ecology of affiliation. Both Blau and Skvoretz take homophily as a primitive assumption in their theories (Blau calls it "salience," Skvoretz calls it "tau-bias") and predict that the heterogeneity of groups will result from this bias and from the opportunity structure afforded by the distribution of people in categories (male vs. female, white vs. African American). Note that this demographic opportunity structure is what provides us with the baseline to which we compare our observed homogeneity (see discussion above and Figures 2-5). The energy distribution principle makes predictions about the deviations from this baseline, or the homophily principle that Blau and Skvoretz use as their unanalyzed assumption.

McPherson's (1983) theory takes a similar tack, and argues that much of the observed homophily in affiliation choices is due to the opportunity structure created by the groups within which these choices are made. Note that this argument is consistent with the energy distribution principle on all fronts: (1) aggregates formed from previously existing collectivities will take less coordination energy than aggregates that form from previously unaffiliated actors; (2) these previously existing affiliations will have been shaped by our principle in the sense that homophilous organizations will find coordination and control of activities easier to manage and will therefore be more likely to survive, and (3) choices made within the opportunity structure of previous affiliation will still display greater homogeneity than chance (as observed by McPherson & Smith-Lovin 1987) because of the cognitive and affective dynamics discussed above as special cases of the principle.

GENERAL ALTERNATIVE THEORIES: SPACE, TIME, AND ENERGY

Under certain circumstances, arguments that rely upon variables mapped on space, time, or energy can be translated into one another, e.g., Stephan (1979). Stephan and McMullin (1981) have suggested that their time-minimization theory in sociology may play a role similar to that of the principle of energy-minimization in physics or cost-minimization in economics. We applaud the efforts of those like Stephan and his colleagues. We have no objection at all to attempts to formulate a theory of social organization around the idea of time-minimization, including in that enterprise the translation of spatial and energy arguments into time arguments. However, we think care should be exercised in making the translations. It should never be assumed that any energy argument, or any space argument, can just automatically be given a time interpretation. The same caution holds for attempts to convert other variables (space, time) into energy variables. When such translations are made, they should be spelled out carefully and with a view to enhancing our ability to systematize and generalize our knowledge. We assume that such care has been taken up to the present time, so our remarks are intended to apply to future attempts. It is entirely possible that a superior theory can be constructed for a wide range of social phenomena based on time-minimization alone. However, it is equally possible that a superior theory could be constructed from energy arguments alone. Our suspicion at the present time is that a superior theory is more likely to be a compound argument employing both time and energy variables.

ENERGY DISTRIBUTION AS A COVERING MECHANISM

As indicated by the discussion above, the energy distribution principle has the advantages that it can predict both (1) the frequency distribution of face-to-face groups and (2) the relative level and form of their homogeneity, relative to a reasonable baseline afforded by the demographic opportunity structure. It acts as a covering mechanism, encompassing a large number of interesting, useful theories that researchers have developed to describe one or more of these phenomena (see Mayhew 1984b:310-12 for a discussion of covering laws). In that sense, it does not refute these other theories. It offers a more general functional form of the relationships that they predict. We argue that these general statements of functional relationships (of the general form "A is a [specified] function of B," where both A and B are variables) are the appropriate goals for scientific theory (Mayhew 1984b:310-11). The specification of so-called causes that discuss unobservable mechanisms are not relevant to scientific discourse *unless* they can be translated into qualifications — further specifications — of relations among observables.

In other words, we have not proposed our energy distribution principle with a view to suggesting that it is either (1) true or (2) superior to other arguments that might be constructed. On the contrary, following Turner (1984), we challenge others to construct better theories, to improve, reject, or reformulate our arguments through research and improved theoretical statement. In the spirit of our belief that a theory should be judged by its correspondence with data and by its fruitfulness for prediction in a wide array of social domains, we

conclude with two applications of our energy distribution principle to substantive areas rather different from those that we have dealt with in this article. We propose these ideas as an encouragement for others to collect different types of data with which to test our ideas. Our ideas relate to the historicity of interactions and to stratification by income.

HISTORICITY

When we observe two people in social interaction, we may find upon inquiry that they have known each other for a long time, that they have a *history* of repeated interactions with one another. Clearly, we can divide interaction sequences (Collins 1967) into two broad categories: (1) those that are repeated interactions between two selected individuals and (2) those that happen for the first time between two selected individuals. The first are customarily conceived — when repeated sufficiently often — as leading to primary group (personal) relations, while the second are relegated to secondary group (impersonal) relations.

When we observe interaction sequences that repeatedly pair the same two individuals, we are seeing the operation of the energy distribution principle. The more frequently two individuals have interacted with one another in the past, the more they share information in general, and the more they share information about one another in particular (Carley 1991). That is, the two individuals are in the process of learning one another's subcultural codes (position language). If the two people occupy nearly identical social positions, the amount of learning of a different subcultural code is at a virtual minimum. If the two people differ on one or more characteristics, then repeated interactions permit them to become familiar with the other's position code, *so that* the amount of energy expended in interaction steadily decreases over time as interactions are repeated. The more the two people learn one another's language, the less energy each needs to expend to communicate with the coordinate action with the other. Thus, regardless of how many interactions we see in a given population, when we partition these into (1) instances of repeated interactions between the same persons and (2) new interactions between two persons, we should observe that a given individual will be devoting most of his or her interaction time to repeated interactions (because these expend less energy per unit of time) and less of his or her interaction time to first-time (new) interactions (because these require greater energy expenditures per unit of time).

In fact, we can say that these episodes of repeated interactions between the same two persons constitute a special case of the "Axiom of Cumulative Inertia" (McGinnis 1968:716): The more often two people have interacted in the past, the greater is the probability that they will interact in the future. This is, of course, because such future interactions expend less energy than future interactions with new (unknown) others.[27] We can therefore develop a whole continuum for the interaction frequencies of each individual, based on the dyads in which she or he participates. For a particular individual A, we can say that A's interaction time will be distributed so that A interacts most often in dyads (triads, tetrads, etc.) containing others with whom A has most often interacted in the past, and

so that A interacts least often in dyads (triads, tetrads, etc.) containing others with whom A has interacted least often in the past. Thus, the energy distribution principle governs each individual's interaction-time distribution *in terms of which other individuals are involved in each time slot.*

This repeated interaction between persons, or *historicity,* emerges as a special case (or consequence) of the energy distribution principle along more than one path. Recall that the law of distance interaction is a special case of the energy distribution principle: people interact most often with others near to them in physical space and less often with those farther away. Furthermore, the movement of people across physical space is governed by the same principle (Zipf 1949). People move about in space (cover physical distance) frequently through short moves and only very infrequently through long moves. So consider each individual's stability-movement distribution. For moves of the individual across space (in the plane) over the times $t_0, t_1, t_2, t_3, \ldots, t_N$, we observe a random walk whose transition probabilities are given by a first order Markov chain: a random walk *biased* so that each move in space is governed by physical distance. That is, each move has a higher probability of being over a small distance than over a long distance. When this process is visualized for an entire population perambulating along random walks in physical distance (in the plane), we find that at t_N each individual has not moved very far from his or her original position at t_0.[28] This means that *most individuals are highly stable in physical space.* And the indicated stability means that the randomly selected individual will be in the same geographic locale most of the time. Hence, each individual will be close to others who also do not venture far from the same locale most of the time. And because people interact most often with those near in physical space, each individual will be in interaction with the *same* other persons most of the time, i.e., those others who also occupy that individual's locale. Accordingly, historicity appears to be, at least in part, a consequence of the energy distribution principle through the law of distance interaction.[29]

STRATIFICATION

An increase in positional homogeneity as a function of group size could clearly come about in more than one way. In our hypothesis 2 and hypothesis 3 we have discussed the pressures toward homogeneity as a result of the energy expenditure required for coordination among group members. We made no predictions about the *content* of the groups, that is, what types of people will be found in groups (aggregates) of size one, two, and so forth. That heterogeneity in social positions would be expected to decline as a function of group size (partially) *through a bias in category composition* is predicted by our energy distribution principle through two lines of status argument.

First, people who are of high status on income tend to be more heterogeneous with respect to most other social characteristics (e.g., occupational position) than are people who are of low status on income.[30] This means that, when summing over all the social distances between them (on occupation, education, etc.), low-status persons are on the average closer to one another in social distance than are high-status persons. Hence, there are fewer (energy) barriers between low-status persons than there are between high-status persons.

The energy distribution principle predicts that the aforementioned size-heterogeneity relation we expect to observe will come about in part because of an increasing proportion of low-status individuals in larger groups. Concretely, relatively small groups (and aggregates of size one) will be disproportionately composed of high-income people; relatively large groups will be disproportionately composed of low-income people. Aggregates of size one are included in this generalization, of course, because high-status persons — being far from most other persons on most social dimensions — will have fewer similar persons to associate with. High-status persons would therefore appear in public alone at a higher base rate than low-status persons.

Second, a status argument can rest on the assumption that *money is an energy converter*. Concretely, in a money economy like that of the U.S., the more money one has, the more energy one can drain from the system (in the form of transport vehicles, gasoline, etc.). Money dissolves energy barriers that would otherwise block access to public places. Money controls differential opportunities for energy expenditures (Galaskiewicz 1979:1348). Accordingly, a deduction from the energy distribution principle is that the distribution of income across persons will be highly skewed such that a relatively small number of individuals will have large amounts of money and a relatively large number of individuals will have small amounts of money. That this is the case is well known.[31] To emphasize this point: the energy distribution principle predicts that any event that involves the expenditure of energy must conform to its shape. Since the distribution of money across persons is a measure of their differential opportunity to expend energy, the income distribution is merely a special case of the energy distribution principle. Economic resources must be expended (on transport, etc.) to go to public places. The lower the income status of people, the more likely they are to pool resources (money, cars, etc.) to go to public places. The higher the income status of people, the less dependent they are on others for economic resources: they are comparatively free to go to public places alone. Because low-status people often must pool resources to venture into public places and high-status people are not so constrained, high-status people should be disproportionately represented in small groups as well as in one-person aggregates; and low-status people should be disproportionately represented in large groups.

Whether from the first or the second line of status argument, we expect that the average status of group members will be a decreasing function of group size (including aggregates of size one). Both status arguments can easily be correct, with each contributing in some degree to the observed outcome. The first status argument refers to energy expended within the group in the process of social interaction itself. The second status argument refers to energy expended to go into public places (on transport, etc.). Thus, each focuses on a different energy event, but both predict the same outcome for group composition as a function of group size. It is important to bear in mind that the first status argument hinges on the assumption that high-status persons are (on the average) farther from one another on the social distance scale than are low-status persons; while the second status argument hinges on the assumption that money is an energy converter (or serves as a mechanism for converting information — the numerical value of each monetary unit — into energy).[32]

Summary

In this article, we have argued that the energy distribution principle can be applied to many domains of concern and can encompass many more specific models. We found empirical support for three predictions about the size and homogeneity of naturally forming face-to-face public groups. The energy distribution principle predicted not just the relationship between size, frequency, and homogeneity of groups, but also the functional form of those relationships. We closed with applications of our general principle to areas about which we do not have data at this point: the interaction histories of persons in groups and the income structure of groups. We hope that these ideas will stir others to apply the energy distribution principle to diverse domains of inquiry.

Notes

1. Jonathan Turner (1984) has suggested that sociologists might do well to abandon attempts to construct *a* theory or *the* theory for each of their respective domains of inquiry. Instead, he suggested, sociologists would benefit from concentrating on the development of *general principles,* principles sufficiently general that they may be applied to more than one domain of concern, e.g., to stratification, to group formation, to bureaucracy, to norm formation. The result of these attempts — if successful — would be a storehouse of general principles, an arsenal of ideas that sociologists could use to tackle almost any problem. Basically, Turner (1980), like Stephan (1979:821), has advocated a return to the strategy of Comte's social physics. See also Stewart (1948). While we do not here advocate abandoning attempts to construct theories about particular phenomena (such as stratification), Turner's general suggestion seems very reasonable to us and the present essay is an attempt to follow his advice.

2. One reader suggests that the energy distribution principle leads to implausible predictions such as cloudy days being more frequent than sunny days and that its scope should be limited only to social events. On the contrary, the principle is widely accepted as describing the operation of physical phenomena (Tribus & McIrvine 1971). It is our challenge here to argue for its applicability as a general, covering principle in the operation of social systems as well. The apparent implausibility of the sunny versus cloudy day prediction stems from the logical flaws in its derivation from the principle. Sunny and cloudy days, as events, do not entail differing amounts of energy; the amount of solar energy reaching the earth's surface during any 24-hour period is roughly constant. Weather events in a particular locale are created by heat transfers that result from upper-level wind and moisture flows. Only from the perspective of *an individual* at a particular place on the earth's surface does weather affect energy received. It is the event itself, not the individual's access to it, to which the energy distribution principle applies.

3. On the role of technology in creating different energy levels in human communities and societies, see Lenski and Lenski (1982) and White (1943). On energy flows created by different technologies, see Cook (1971), Kemp (1971), and Rappaport (1971).

4. Mayhew (1983:176-90) outlined a frame of reference for studying the impact of sharp shifts in energy level on human social organizations (more generally, see Kreps 1984 and references therein).

5. For example, one reader suggests that the principle would predict that babies would be more common than middle-aged people because they expend less energy. This prediction is not consistent with the principle, because the energy distribution principle applies to events rather than people.

6. This outcome has been observed in several studies (Cohen 1971; James 1951, 1953).

7. As Elijah Anderson (1990) describes vividly, people also can expend significant amounts of energy to *avoid* contact with one another. The most straightforward way to study this

phenomenon using the energy distribution principle would be to define a domain of avoidance events that could then be hypothesized to vary in frequency with the strength and difficulty (energy expenditure) of the avoidance. How would we expect avoidance to affect the energy expenditure of groups and its relationship to group size? If avoidance is a more salient issue for single individuals in public alone, the energy expenditure devoted to avoidance would reduce the effectiveness of group size as a measure of energy expenditure and would work against our hypotheses (discussed below). If avoidance energy expenditures increase with the boundary maintenance problems encountered by larger groups, then they will strengthen our use of group size as an indicator of energy expenditure. Clearly, we argue for the latter interpretation, which is simply a special case of the principle.

8. As we elaborate below, a multiplicative increase does not have to literally double or triple in effect with an addition to the group. What we argue is that the energy increase follows a quadratic function, specified later in equation 2.

9. A reader argues that increasing the division of labor would reverse this relationship. We argue that, on the contrary, the division of labor is at least partly a response to the energy distribution principle. As systems increase in size, the amount of energy expended in coordination would reach impossible levels without an increase in structure (Mayhew et al. 1972). By structure we mean some arrangement whereby not all possible connections in the system will occur. Whether this increase in structure takes the form of functional differentiation (Smith 1791), hierarchy (Mayhew 1983; Williamson 1975), or some other adaptation will be dictated by other considerations.

10. Notice that the domain must be defined to include all relevant events, rather than those within a limited energy or size range. For example, we would want to include all storms rather than just hurricanes, since hurricanes are storms reaching a certain minimum energy expenditure. If we study transportation events, we would not want to restrict our attention to buses or trains, since they are specifically designed for *mass* transit; instead, we would look at all instances of transportation. Then we would certainly find that people traveling alone (in cars or by foot) would be the most numerous category. Households are an interesting case. The modal household appears to hold two people (cf. Davis & Smith 1990). Frequencies above two show the pattern the principle would predict, but one-person households are slightly less common than two-person households. Is this a disconfirmation? First, we have to ask whether a household is an event. The more appropriate unit of analysis for testing the principle would be the actually occurring aggregate, rather than the officially defined household unit. In other words, a household is an event only when all its members are actually copresent and interacting. By this measure, the number of people who are "home alone" when contacted would doubtless push the number of one-person aggregates upward. In addition, we find that official statistics and common surveys define out of their population or fail to find many individuals who live as singletons: the institutionalized, the homeless, members of the armed forces, students, etc. If such cases were included, the modal category would certainly be one.

11. We will not be making direct tests of the relationship between physical distance and social interaction frequencies. Distance is discussed here as an item subsumed under the energy distribution principle because it enters into the interpretation of our results.

12. For references to research supporting this proposition, see the materials cited in Mayhew and Levinger (1976:93-94).

13. We will not be making a direct test of the impact of language on the frequency of social contact (but see Carley 1991). However, language differences at the level of subculture (including differences in dialects) enter into our reasoning about social distance. Hence, it is important to note that interaction frequencies across and within language groups are also a special case of our energy distribution principle.

14. We do not deny that people may communicate through "artificial" languages and through technologically enhanced channels (Kaufer & Carley 1993). However, since most artificial languages and augmented channels are in fact peculiar to certain subcultures, our subsequent discussion of sexual, ethnic, racial, occupational, and income subcultures permits us to include these artificial communication codes and channels in our arguments.

15. A narrower version of this hypothesis might be called the status distance interaction hypothesis. It claims that the frequency with which people enter into face-to-face association

(contact, interaction) is inversely related to the status distance between them. Because it refers to rank differences only, rather than to a more general difference in social position, this is a much narrower hypothesis than the social distance hypothesis discussed in the text. It becomes narrower still if status is further interpreted — as it usually is — to mean only rank on the vertical dimension of social approval (i.e., prestige, esteem, or social honor). Because some segments of the literature have managed to impose this narrower interpretation (Ellis 1956), a correction is required. In the present article, status refers to any *ranked social position* on any of the commonly used dimensions of vertical differentiation, especially including wealth, power, and prestige, but also including general indicators of these characteristics where appropriate (e.g., income, years of school completed). We have taken the time to specify this hypothesis because, while it is narrower than the social distance hypothesis, it may be that rank on vertical differentiation has more powerful effects than other kinds of social distance. If that is so, the status distance hypothesis is an extremely important special case of the social distance hypothesis, one that we need to keep clearly in mind when considering the relative explanatory power of different dimensions.

16. An objectivist definition of social distance is adopted here (cf. Blau 1977). In this external view, social distance is imposed (defined) by the outside observer. A subjectivist definition of social distance conceives of this dimension from the point of view of the individual observed (cf. Bogardus 1959). That is, social distance is given by the attitudes or feelings of individuals toward one another, and, consequently, may vary considerably from one individual to another. We do not consider the subjectivist approach here.

17. See the overview in Marsden and Laumann (1984:279-81). Extension of this idea to continuous variables appears in McPherson (1983) and McPherson, Popielarz, and Drobnic (1992).

18. This may seem much more obvious to the reader for some kinds of subcultural differences than for others, especially when differences are strong enough to take on the character of dialects within the same natural language. But we are arguing much more than this. We are saying that each position has its own subcultural code, that code being a variant on the common natural language spoken in a given society. According to Mannheim (1980), individuals who are classed together in the same concrete group share a basic stock of experiential contents. This could be true, of course, without requiring that these experiences constitute a distinct subcultural code. However, if we adopt Marx's (1939:80) view that language and thought are the same thing, then we would almost certainly require that people who occupy different positions do speak in different codes, even when they are using the same words in a natural language. To the extent that people occupy similar social positions, they experience many things in the same way *by virtue of that occupancy*. This increases the likelihood of contact or communication between them, because they speak the same language (think the same things, in Marx's view). The "contents of their consciousness are similar" (Vierkandt 1923:261-62). In stressing these similarities for people who occupy the same (or structurally equivalent) positions, we are not referring to the "pressures toward uniformity" in beliefs, actions, etc. that are commonly argued for people throughout the entire system (Allardt 1964). On the contrary, we are saying that this uniformity of experience, thought, and language is confined to particular positions within society. Those who remain skeptical that most social positions will have their own subcultural codes would do well to consult Bernstein (1960) and Schatzman and Strauss (1955) before prematurely rejecting our interpretation. For a more recent statement that appears to agree exactly with our argument, see McNall (1985:23). Simmel (1911:14) argued that people who are different from one another have difficulty in understanding one another.

19. One might raise the question of whether the fact that many groups arrived at the observational scene in cars could have constrained the size distribution; most cars can hold only two to four people comfortably. More generally, our observations of group size undoubtedly constrained by a large number of external social factors such as the size of households, the physical capacity of the setting to accommodate large groups, etc. We are encouraged by two features of this problem. First, in our empirical analyses we see strong relationships between group size and frequency (as predicted by our hypothesis hypothesis 1) both above and below the threshold that would be physically imposed by the capacity of the typical car (i.e., in the range 1-5 and in the range 5-11). Therefore, we are quite sure that the

strength of our relationship is not created by the transportation constraint. Second, the tests of hypothesis 2 and hypothesis 3 are not affected by any external constraint on group size. In general, we believe that the social constraints have evolved within the constraints of the energy distribution principle. We suspect that if people typically traveled in larger groups, transportation devices and public places would be designed to accommodate them (just as station wagons were more common when the average American family was larger). At any given point in time, however, these constraints shape the energy expenditure entailed by a given group size (e.g., it is more difficult to move a large group to a public event than a small group).

20. With respect to errors in observation, we assume that these will occur for any observation procedure that might be devised. Further, we doubt that two different observers — given the same location and time of day — would note exactly the same number of groups. We can only hope that over a large number of studies conducted by a large number of different observers, the errors will be small in number and random. Note that any tendency to notice systematically larger groups or groups of unusual composition will tend to work against our hypotheses. Therefore, we hope that our tests are conservative. All the data in the present study were collected by the first author.

21. One reviewer has suggested that we compare our size distribution of face-to-face groups with size distributions from other empirical settings (e.g., towns or firms). Such a comparison would bias the hypothesis test in our favor, since we believe that both empirical phenomena are generated by the same covering principle: the energy distribution principle. We prefer to use a baseline model that is clear in its assumptions (Mayhew 1984a) and that can be evaluated rigorously in terms of its accuracy in describing the data. While observing similarity of empirical distributions would pose an interesting question for future research (one that we trust would be answered through application of the energy distribution principle), such comparisons would be ad hoc in the sense that there would be no clear statement of the analytic principles that would generate such a distribution.

22. Negative binomial regression analysis can be regarded as a generalization of Poisson regression analysis, which parameterizes overdispersion in count data (see Barron 1992). We estimate the parameters of the models with King's COUNT procedure (King 1989a, 1989b).

23. We use pooled observations from our sweeps through public places rather than U.S. Census data to estimate the baseline proportions for several reasons. First, it is impossible to locate the boundaries of the population from which the face-to-face groups are being drawn. For example, Columbia's public may be drawn primarily from the city, its county (Richland County), or the two-county SMSA that constitutes its economic base (Richland and Lexington counties). The racial proportions would vary dramatically with the catchment area we assumed (from 39.7% to 38.2% to 28.5% African American, respectively). Similarly, Winnsboro is 50.5% African American, while Fairfield County is 57.6% African American. A second problem with using census figures for a baseline is the fact that the probability of a person appearing in public differs by race and sex. In both towns, African Americans are less likely to appear in public than their census populations would lead us to expect. In Winnsboro, 46.3% of people who appeared in our sweeps were African American; in Columbia, 31.5% were African American. Similarly, women are somewhat less likely than men to appear in public; women made up 46.3% and 54.2% of the publicly observed population in Winnsboro and Columbia respectively. These figures represent less than their census representation in both cases. We argue that social forces possibly related to energy distribution, such as residential distribution, availability of transportation, etc., create differences in public appearances.

24. Blau's terminology is in some instances different from our own; we mention these differences in order to avoid confusion. Blau uses the term *group* to include not only the traditional instance of people who are bound together in interaction, but also for social categories as traditionally defined (people with a common social position or characteristic, but without the implication that these individuals are in interaction with one another). Our use of *group*, especially face-to-face group, means exactly and only instances of people who are in direct face-to-face social contact with each other. At times we introduce the term *aggregate* to indicate that we are considering not only the size distribution of face-to-face groups, but also — added to these — aggregates of size one (lone individuals). Blau uses the term *parameter* for what we call a social dimension (age, sex, race, ethnicity). He distinguishes graduated parameters (age, income) from nominal parameters (sex, language group). Since graduated

parameters include our dimensions of vertical differentiation (status), they are quite important (in our view as well as Blau's). It is worth mentioning that our theoretical concerns are quite different from those addressed by Blau, although they overlap in some instances. Blau's theory is confined to dyadic contacts, whereas our concern is with face-to-face groups of all sizes, not just dyads. Thus, while most of Blau's theoretical concerns overlap ours, we have additional concerns (groups larger than two persons and, in some instances, aggregates of size one). Finally, our concerns are only with face-to-face interactions, whereas many of the contacts that can be subsumed under Blau's theory need not be restricted to face-to-face interaction.

25. We, of course, recognize that experiences in the workplace will differentiate high-income (on average) men from their lower-income (on average) wives. Also, power differences within the family created by relative income contributions of husband and wife may affect interaction patterns and create variations in gender-based subcultural codes.

26. Note that the major exception to this argument would be the army base at Fort Jackson in Columbia, South Carolina. We hypothesize that observations there might reverse the finding.

27. Simmel (1911:14) argued that we have less difficulty understanding a member of our own family (someone with whom we interact often) than we have understanding a stranger (someone with whom we have no past history of interaction).

28. Since a model of movement in the plane that permits each individual to select a direction of movement from each of 360 degrees, and then a distance to move in that selected direction, may seem unduly complicated for clear illustration in this simple discussion, more elementary models for such a random walk may be preferred. Following Eigen and Winkler (1975:50-52), we might construct such a walk on a board for a glass bead game, e.g., a checkerboard. Here, the individual is positioned in the middle of the checkerboard at t_0 and has five alternative transitions between $t - 1$ and t. She or he can (1) remain in her or his present position (the highest probability from the law of distance-interaction) or (2) move in any of four directions across the checkerboard, with the number of squares crossed in any move having transition probabilities that are a decreasing function of their distance from the individual's position at $t - 1$. A simpler model is obtained from a board with equilateral triangles; a more complex one from a board consisting of equilateral hexagons. From hexagons, we can move on to the more elaborate illustration with 360 rays out from the center of the plane.

29. Simmel (1898:667) said that the first thing determining the unity of human groups is spatial proximity of persons.

30. The argument is not that high status persons are farther apart on *all* position dimensions. (Indeed, there is some evidence from network studies that higher-status people have more system-spanning ties than low-status people.) Rather, they are farther apart on some *important* position dimensions. That some social categories that are low on income tend on the average to be more homogeneous on other social characteristics (such as occupation) has long been recognized. In the U.S., the case of African American females is particularly striking for occupational homogeneity (Blau 1977; Gross 1968). Incidentally, the dimension that determines status in this discussion — income — would itself be expected to show a greater dispersion of high-status persons: high-income persons (those above the median income) will be farther apart on income than will low-income persons (those below the median income).

31. Most income distributions approximate the lognormal distribution and therefore show a high degree of skew with a small number of high incomes and a large number of small incomes. See Champernowne (1973) and Zipf (1949) in general, and Atchinson and Brown (1957:107-20) in particular.

32. On the relationship between information and energy in this regard, see Tribus and McIrvine (1971). On the nature of money as a generalized resource in highly developed money economies, see Simmel (1900). On the conditions that give rise to money economies, see Elias (1939:61).

References

Allardt, Erik. 1964. "A Theory on Solidarity and Legitimacy Conflicts." *Transactions of the Westermarck Society* 10:78-96.

Anderson, Elijah. 1990. *Streetwise: Race, Class, and Change in an Urban Community*. University of Chicago Press.

Atchinson, J., and J.A.C. Brown. 1957. *The Lognormal Distribution*. Cambridge University Press.

Barron, David. 1992. "The Analysis of Count Data: Overdispersion and Autocorrelation." Pp. 179-230 in *Sociological Methodology 1992*, edited by Peter V. Marsden. Basil Blackwell.

Bernstein, Basil. 1960. "Language and Social Class." *British Journal of Sociology*. 11:271-76.

Blau, Peter M. 1977. *Inequality and Heterogeneity*. Free Press.

Blau, Peter M., and Joseph E. Schwartz. 1984. *Crosscutting Social Circles: Testing a Macrostructural Theory of Intergroup Relations*. Academic Press.

Bogardus, Emory S. 1959. *Social Distance*. Antioch Press.

Carley, Kathleen. 1991. "A Theory of Group Stability." *American Sociological Review* 56:331-54.

Carley, Kathleen. 1993. "Coding Choices for Textual Analysis: A Comparison of Content Analysis and Map Analysis." In *Sociological Methodology 1993*, edited by Peter V. Marsden. Basil Blackwell.

Carley, Kathleen, and Michael Palmquist. 1992. "Extracting, Representing, and Analyzing Mental Models." *Social Forces* 70:601-36.

Champernowne, David Gawen. 1973. *The Distribution of Income between Persons*. Cambridge University Press.

Cohen, Joel E. 1971. *Casual Groups of Monkeys and Men*. Harvard University Press.

Coleman, J.S. 1962. "Comment on Harrison White, 'Chance Models of Systems of Casual Groups.'" *Sociometry* 25:172-76.

Coleman, J.S., and J. James. 1961. "The Equilibrium Size Distribution of Freely-Forming Groups." *Sociometry* 24:36-45.

Collins, Orvis. 1967. "The Elemental Processes of Social Interaction." *Human Organization* 26:6-21.

Cook, Earl. 1971. "The Flow of Energy in an Industrial Society." *Scientific American* 225:134-44.

Davis, James Allen, and Tom W. Smith. 1990. *General Social Surveys 1972-90: Cumulative Codebook*. National Opinion Research Center.

Durkheim, Emile. 1893. *De la Division du travail social*. Paris: Felix Alcan.

Eigen, Manfred, and Ruthild Winkler. 1975. *Das Spiel: Naturgesetze steuern den Zufall*. Munich, Ger.: R. Piper.

Elias, Norbert. 1939. *Über den Prozess der Zivilisation*, vol. 2. Haus zum Falken.

Ellis, Robert A. 1956. "Social Status and Social Distance." *Sociology and Social Research* 40:240-46.

Eysenck, Michael W. 1984. *A Handbook of Cognitive Psychology*. Erlbaum.

Festinger, Leon. 1954. "Theory of Social Comparison Processes." *Human Relations* 7:117-40.

Frazier, Kendrick. 1982. *Our Turbulent Sun*. Prentice-Hall.

Freeman, Linton C. 1984. "Turning a Profit from Mathematics: The Case of Social Networks." *Journal of Mathematical Sociology* 10:343-60.

Galaskiewicz, Joseph. 1979. "Structures of Community Organizational Networks." *Social Forces* 57:1346-64.

Goodman, Lee A. 1964. "Mathematical Methods for the Study of Systems of Groups." *American Journal of Sociology* 70:170-92.

Green, L. C. 1985. *The Town of Winnsboro, South Carolina: 1983-1984 Annual Report*. Winnsboro: Office of the Town Manager.

Gross, Edward. 1968. "Plus Ça Change . . .? The Sexual Structure of Occupations over Time." *Social Problems.* 16:198-208.

Hawley, Amos. 1971. *Urban Society: An Ecological Approach.* Roland.

Hosmer, D.W., and S. Lemeshow. 1989. *Applied Logistic Regression.* Wiley.

James, John. 1951. "A Preliminary Study of the Size Determinant in Small Groups." *American Sociological Review* 16:474-77.

____. 1953. "The Distribution of Free-Forming Small Group Size." *American Sociological Review* 18:569-70.

Kaufer, David S., and Kathleen M. Carley. 1993. *Communication at a Distance: The Influence of Print on Socio-cultural Organization and Change.* Erlbaum.

Kemp, William E. 1971. "The Flow of Energy in a Hunting Society." *Scientific American* 225:104-15.

King, Gary. 1989a. "Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator." *American Journal of Political Science* 33:762-84.

____. 1989b. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference.* Cambridge Press.

Kreps, Gary A. 1984. "Sociological Inquiry and Disaster Research." Pp. 309-30 in *Annual Review of Sociology,* edited by Ralph H. Turner and James F. Short. Annual Reviews.

Lenski, Gerhard E., and Jean Lenski. 1982. *Human Societies: An Introduction to Macrosociology,* 4th ed. McGraw-Hill.

Mannheim, Karl. 1980. *Structuren des Denkens.* Frankfurt-am-Main, Ger.: Suhrkamp.

Marsden, Peter V., and Edward O. Laumann. 1984. "Mathematical Ideas in Social Structural Analysis." *Journal of Mathematical Sociology* 10:271-94.

Marx, Karl. 1939. *Grundrisse der Kritik der politischen Ökonomie (Rohentwurf 1857-1858).* Moscow, Russ.: Verlag für fremdsprache Literatur.

Mayhew, Bruce. 1980. "Structuralism vs. Individualism: Part I, Shadowboxing in the Dark." *Social Forces* 59:335-75.

____. 1981. "Structuralism vs. Individualism: Part II, Ideological and Other Obfuscations." *Social Forces* 59:627-48.

____. 1983. "Hierarchical Differentiation in Imperatively Coordinated Associations." Pp. 153-229 in *Research in the Sociology of Organizations,* edited by Samuel Bacharach. JAI Press.

____. 1984a. "Baseline Models of Sociological Phenomena." *Journal of Mathematical Sociology* 9:259-81.

____. 1984b. "Chance and Necessity in Sociological Theory." *Journal of Mathematical Sociology* 9:305-39.

Mayhew, Bruce, and Robert L. Levinger. 1976. "Size and the Density of Interaction in Human Groups." *American Journal of Sociology* 82:86-110.

Mayhew, Bruce, Robert L. Levinger, J. Miller McPherson, and Thomas F. James. 1972. "System Size and Structural Differentiation in Formal Organizations: A Baseline Generator for Two Major Theoretical Propositions." *American Sociological Review* 37:629-633.

McGinnis, Robert. 1968. "A Stochastic Model of Social Mobility." *American Sociological Review* 33:712-22.

McNall, Scott G. 1985. "Prisoners of Language." *Contemporary Sociology* 14:23-26.

McPherson, J. Miller. 1982. "Hypernetwork Sampling: Duality and Differentiation among Voluntary Associations." *Social Networks* 3:225-49.

____. 1983. "An Ecology of Affiliation." *American Sociological Review* 48:519-32.

____. 1990. "Evolution in Communities of Voluntary Organizations." Pp. 224-25 in *Organizational Evolution: New Directions,* edited by Sage. Jitendra V. Singh.

McPherson, J. Miller, Pamela A. Popielarz, and Sonja Drobnic. 1992. "Social Networks and Organizational Dynamics." *American Sociological Review* 57:153-70.

McPherson, J. Miller, and Thomas Rotolo. 1995. "Measuring the Composition of Voluntary Groups: A Multitrait-Multimethod Analysis." *Social Forces* 73:1097-1116.

McPherson, J. Miller, and Lynn Smith-Lovin. 1982. "Women and Weak Ties: Differences by Sex in the Size of Voluntary Organizations." *American Journal of Sociology* 87:883-904.

____. 1986. "Sex Segregation in Voluntary Associations." *American Sociological Review* 51:61-79.

____. 1987. "Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups." *American Sociological Review* 52:370-79.

Rappaport, Roy A. 1971. "The Flow of Energy in Agricultural Society." *Scientific American* 225:116-32.

SAS/STAT User's Guide. 1989. Version 6, 4th ed., Vol. 2, SAS Institute.

Schatzman, Leonard, and Anselm Strauss. 1955. "Social Class and Modes of Communication." *American Journal of Sociology* 60:329-79.

Schneider, Stephen H., and Randi Londer. 1984. *The Coevolution of Climate and Life.* Sierra Club Books.

Simmel, Georg. 1894. "Le Problème de la Sociologie." *Revue de Métaphysique et de Morale* 2:497-504.

____. 1898. "The Persistence of Social Groups." *American Journal of Sociology* 3:662-98.

____. 1900. *Philosophie der Geldes.* Leipzig, Ger.: Duncker & Humblot.

____. 1911. *Philosophische Kultur: Gesammelte Essaise.* Leipzig, Ger.: W. Klinkhardt.

Skvoretz, John. 1983. "Salience, Heterogeneity and Consolidation of Parameters." *American Sociological Review* 48:360-75.

Smith, Adam. 1791. *An Inquiry into the Nature and Causes of the Wealth of Nations.* 6th ed. London: A. Strahan & T. Cadell.

Stephan, Edward G. 1979. "Derivation of Some Social-Demographic Regularities from the Theory of Time-Minimization." *Social Forces* 57:812-23.

Stephan, Edward G., and Douglas R. McMullin. 1981. "The Historical Distribution of County Seats in the United States: A Review, Critique, and Test of Time-Minimization Theory." *American Sociological Review* 46:907-16.

Stewart, John Q. 1948. "Demographic Gravitation: Evidence and Applications." *Sociometry* 11:31-58.

Tribus, M., and E.C. McIrvine. 1971. "Energy and Information." *Scientific American* 255:179-88.

Turner, Jonathan. 1980. "Toward a Social Physics: Reducing Sociology's Theoretical Inhibitions." *Humboldt Journal of Social Relations* 7:140-55.

____. 1984. *Societal Stratification.* Columbia University Press.

U.S. Census of the Population. 1980. *Volume 1: Characteristics of the Population.* Chapter B. General Population Characteristics. Part 42: South Carolina. Department of Commerce.

Vierkandt, Alfred. 1923. *Gesellschaftslehre.* Stuttgart, Ger.:F. Enke.

White, Harrison. 1962. "Chance Models of Systems of Casual Groups." *Sociometry* 25:153-72.

White, Leslie. 1943. "Energy and the Evolution of Culture." *American Anthropologist* 45:335-56.

Williamson, Oliver E. 1975. *Markets and Hierarchies: Analysis of Antitrust Impliations.* Free Press.

Zipf, George Kingsley. 1949. *Human Behavior and the Principle of Least Effort.* Addison-Wesley.

Emancipating Sociology: Postmodernism and Mainstream Sociological Practice*

JOEL SMITH, *Duke University*

Abstract

Using analyses of mass communication and research methodology, this article assesses the validity of the postmodernist critique for sociology. These analyses locate various weaknesses and inconsistencies in the critique. Three hypotheses are proposed to account for the tolerance the critique receives from many sociologists despite its recognized inadequacies. Some implications for the discipline of the factors involved in those hypotheses are discussed. The analysis concludes that sociologists are unduly tolerant of and responsive to such critiques.

Mainstream American sociology has long been a target of critical analyses — positivism, neo-Marxism, critical theory, ethnomethodology. The charges vary — sociology is not, cannot, and should not be value-free; it is inadequately scientific; it pretends to be scientific when, in fact, the relation of subjects to investigators makes that impossible; claims to objectivity are false; it should be explicitly political but is not; and so on. Postmodernism is the most recent of these critiques. Despite being of little interest to many American sociologists[1], it is inspiring a burgeoning sociological literature whose eventual impact on the discipline remains to be seen. It could be drastic. The unfettered observation, analysis, and interpretation that postmodernism has encouraged in the humanities may appeal to sociologists concerned with the historical specificity of and reflexive relationship with their material. That may account for the growing use of postmodernist rhetoric in sociological discourse.[2]

The critique draws upon the philosophy of language and science to assert that communication and perception must be flawed and unreliable. This implies

* *This is a drastic revision of an article prepared for a panel on Philosophical Foundations of Sociological Knowledge and Applied Sociology at the 31st Congress of the International Institute of Sociology, 21-25 June 1993. I have had the benefit of responses from William H. Form, Kenneth Land, Seymour Martin Lipset, John O'Neill, Herbert L. Smith, Edward Tiryakian, John Wilson, David Zaret, and three unusually patient and supportive anonymous referees. They identified numerous deficiencies, many of which I shall never be able to address adequately. However, I can and do thank them for their support and willingness to help. Direct correspondence to Joel Smith, Department of Sociology, Duke University, Durham, NC.*

that empirical inquiry is meaningful only to the investigator. If so, it would make no sense to share results because no one could be sure how others would take them. Research would be a selfish, ego-centered pursuit. If this is the prospect, sociologists need to assess the legitimacy and implications of the postmodernist critique for sociology. I do so from the perspective of my own work because a key postmodernist premise — it is impossible to validate interpretations of communication — and a key derivation from it — all methods of inquiry are warranted — appear to agree with positions I have taken. In this article I evaluate parallels between postmodernist analysis and my positions on (1) communication and (2) methodology. I also (3) propose several hypotheses that may account for why some sociologists have embraced and others tolerate the criticism and (4) discuss the implications for the discipline of the factors involved.

Postmodernism

The postmodernist analysis claims that there have been fundamental societal shifts resulting from transformations implicit in capitalism (see Smart 1992:141-82; 1993). Adherents vary on several matters that are not pertinent here[3], but basically they argue that the conditions of modern life altered as the consumerism fostered by capitalism and spread by modern media[4] — their role and weight is debated — transformed people's relations to one another and to their environments (e.g., virtual reality). This process, they claim, has resulted in the replacement of reality with various fictitious realities that cater to and preempt the attention of diverse new groups that transcend older social arrangements. The diversity of interests and activities separate the groups from each other. Change is now so incessant and rapid that there is no opportunity for a new societal integration. Consequently, for Baudrillard (1983:18), the possibility or desirability of social theory and the existence of "the social" are questionable (see Smart 1993:53).

Postmodernism is most widely espoused by scholars in the humanities (e.g., literary criticism, comparative literature, architecture), but its intellectual roots are in Nietzsche's critique of modernism. It has evolved through Heidegger to Barthes and Foucault, and uses "Wittgenstein and Gadamer for philosophical inspiration and . . . academic respectability" (Bauman 1992:40). Its assertions provide the basis for as radical a critique of the epistemology and ontology of positivistic social science as ever has been mounted. Rosenau predicts (1992:171), in understated prose, that, were skeptical postmodernism[5] to prevail, it probably would be mainstream empirical sociology's death knell.[6]

Lyotard (1984:79-82, 1989:7-10; also see Smart 1992:175-76) distinguishes three different connotations of the term, but is confident that there has been a major shift that is consequential for scholarship and knowledge (1984:3). Critics of postmodernism find it difficult to define (see Crook 1991:150; Bernstein, 1992:199-200; Rosenau 1992:7-19; Smart 1992:169-70). Rosenau notes that adherents disagree on its specifications, even debating the hyphen between "post" and "modern." Most writers describe postmodernism in terms of what it rejects. Aronowitz (1988) says it is "marked by the renunciation of foundation-

al thought, of rules governing art, and of the ideological 'master discourses' liberalism and Marxism" (46). Bauman (1992) says it is a reaction to a totalizing idea of reason, but not to rationality. The discussions converge on a conception of postmodernism as an antifoundational, antijustificatory, antirepresentational, critical deconstruction of tradition.[7] Rather than try to span the many differences among postmodernists with a definition, I use the term in this sense because it underlies their commentary on empirical social science.

Mass Communication

Postmodernists challenge the representative function of language (or in some cases anything) because connections between words and their referents are arbitrary and imprecise. Some claim that there is no social reality other than symbolic representations.[8] If so, anything perceivable is a potential symbol, and, hence, a text whose meaning, they argue, is indeterminate for several reasons.

1. Some texts were not produced intentionally, so their meanings only come from their perceivers.

2. The meanings of some intentionally produced texts never were clear because of various possible problems in their production.

3. Regardless of whether reason 1 or reason 2 applies, meanings at text reception can only be revealed in reports produced afterward; their connection to initial meanings is subject to the same uncertainty.

Consensual meanings, then, are unlikely or transitory because, other than by force, there is no way to establish and preserve them, particularly as what Lyotard calls the "metanarratives" of the past lose their force as computers change the nature of language and knowledge (1984:9-17,60-64,75-77). In view of these uncertainties, the possibility of communication and its role in ordering social behavior are always in doubt.[9] Indeed, some skeptics argue that "the social" may not exist and that society is a concept contrived to justify actions (e.g., training, coercion) that render people more controllable.[10] If so, sociology is a political project, not a scholarly discipline (Bogard [1990, 1992], however, suggests that Baudrillard and Lyotard simply consider sociology irrelevant). Aside from its political aspects, the issue of whether somewhat orderly and integrated social life is possible hinges on the issue of the meaning of symbols and the conclusions regarding consensual meanings and what follows from them.

Even though postmodernists may differ on whether a text can be anything (see Rosenau 1992:8), they agree that the most obvious texts are products of what most people consider communication — novels, poems, paintings, films, sitcoms, speeches, remarks, and the like. Inasmuch as the critique is particularly relevant to analyses of mass communication and mass media, it intersects with my interests. Long before these arguments were directed to the social sciences, students in my seminar on mass communication were questioning the precise nature of communication. In light of claims that the media both inform and manipulate, they insisted on the need to distinguish communication from influence. They argued that successful communication occurs only if intentional-

ly produced messages are received by targeted receivers who understand them as intended (see Smith, Bealer & Sim 1962:12). It follows that the criteria for successful communication must include a would-be communicator behaving as specified and receivers with mental states congruent with the intended meaning as a result of the action. Whether outcomes satisfy senders' desires to influence (or, in some cases, receivers' desires to be influenced) is irrelevant.

This concept of communication raises precisely the same problems noted by postmodernists.[11] To determine if communication is occurring requires knowing whom senders intended to reach, what meanings they intended them to have, who received these meanings, the meanings that intended receivers had after the act, and whether they were acquired from the act of reception. It is clear that several conditions make it impossible to access the data needed to be certain whether there is equivalence, and, if there is, whether it resulted from communication. It also is clear that no matter what evidence is adduced, there always will be uncertainty because, short of having reliable non-communication-based measures (e.g., brain activity) of what is happening at every moment with each participant for an indeterminate time span during which both learning and forgetting can occur, the only data on which to decide are reports subject to the same sources of inaccuracy.

Because human social life is so thoroughly involved with and dependent on communication, but neither subjects' reports nor analysts' interpretations of meaning can be validated, postmodernists argue that a social science that requires understanding actors' meanings is impossible. Several observations may be made in this regard. To accept them requires assuming the reality of everyday life, but this seems reasonable so long as postmodernists address others.

(1) Most sociologists assume that they interact with others, can observe others interacting, and that behavior called communication is one mode of contact. Because communication is fraught with error and experienced as risky and difficult, rules to address these difficulties develop in any group whose survival requires members to achieve and maintain shared understanding. The rules do not assure communication, but, in tandem with the bonds that lead people to attend to one another, they hold people together while they try to resolve the problems in sharing meanings. Moreover, motives that lead to an effort to communicate can be satisfied even if meanings are not shared precisely. Pure communication — completely congruent understandings as a result of intentionally transmitted messages — expedites but is not necessary for compatible social action. Most ongoing social relations, the requisite of social organization, can accommodate poorly articulated actions based on dissimilar understandings. Not only do most social relations *not* require perfect communication, this kind of communication could produce conflict.

(2) People are able to accept and take systems for granted in the same way. In analyzing the relation of social action to society, Schutz (1963) uses the postal service to show how trust facilitates the integrated, ongoing relations among strangers that make society possible. Even taken as a text, most people who use (i.e., readers) or work for the postal service (i.e., producers) "read" it similarly. Freedom to differ need not lead to disorder and unpredictability.

(3) Postal service also exemplifies the reality of social structure as a complex of persisting arrangements that transcend time, place, and individuals. There are many other cases of orderly social behavior based on assumptions (e.g., following the rules of the road, queuing up, paying taxes). Without immediate coercion and by relying on socialization through communication, activity among strangers is adequately coordinated and regularized to provide predictability in group life.

(4) The preceding observations imply that people are able to accept the apparent, the relative, the approximate, the taken for granted as real, a capacity captured in Thomas's (1951) famous dictum that "If men define situations as real, they are real in their consequences" (81). The inescapable uncertainties in everyday life that go with the mediation of people's access to reality is a crucial element in the postmodernist critique. However, mediation by the impedimenta of communication is not new.[12] Language alone always has mediated humans' relation to reality (see McLuhan 1964:81-84). Scientists proceed on assumptions and research accommodates indeterminacy — a concept originating in physics, not in social science. Postmodernists would deny social scientists the right to deal with uncertainty in the same way that most people, postmodernists included, do.

(5) Structured relations have effects without communication. The media, themselves, provide examples. New media have such characteristic effects as new consumption patterns for existing media, rescheduling and restructuring public events to meet media needs and schedules, and rearranging personal schedules to fit media schedules. Marshall McLuhan, an icon of many postmodernists, Harold Innis (1964), and others on whom McLuhan drew, note that, rather than destroy social organization, new media reshape it (see Meyrowitz 1985; Zaret 1992).

The assumptions that what is experienced as social reality exists and that people are able to act in an orderly manner on the basis of assumptions despite the uncertainty introduced by language run counter to central points in the postmodernist critique. Although many would demur, postmodernists' actions suggest that they accept these assumptions. It also is not clear that media-mediated reality creates fundamentally different problems from those that people always have faced in symbolically mediated environments. Agreement that an irreducible uncertainty pervades communication does not require accepting the conclusions that postmodernists reach. If people can live effectively with imperfect communication in mediated environments, sociological research with uncertainty should be possible. However, the inescapable problems attending the mediation of reality by symbols, and by reflexivity, are forceful reminders of its limits.

Methodology

The postmodernist analysis of communication undergirds its radical position on social research. "Some of the most important consequences of the postmodern revision of text, author, and reader are methodological in nature" (Rosenau 1992:40). In contrast to conventional social scientists, who "seek to isolate

elements, specify relationships, and formulate a synthesis . . . (postmodernists) offer indeterminacy rather than determinism, diversity rather than unity, difference rather than synthesis, complexity rather than simplification. They look to the unique rather than the general, to intertextual relations rather than causality, and to the unrepeatable rather than the re-occurring, the habitual, or the routine" (8). If this view prevails and if sociology is to continue to be relevant, it must "break its links with the ontological and epistemological premises of modernity . . . by accepting . . . the self-containment and self-grounding of the production and reproduction of meanings" (Bauman 1992:26). That would mean that sociologists could no longer treat social reality as an object and analyze data as if they shared concepts with consensual meanings. In turn, it would require fundamental changes in procedures and substance.

If, as postmodernists claim, it is impossible to validate any interpretation of a text, none are demonstrably right or wrong and all are equally plausible. Consequently, for postmodernists "social science becomes a more subjective and humble enterprise. . . . Truth gives way to tentativeness. Confidence in emotion replaces efforts at impartial observation. Relativism is preferred to objectivity, fragmentation to totalization" (Rosenau 1992:8). Moreover, those who argue that there is only representation and no reality admire (41) and even encourage the production of numerous, possibly inconsistent, interpretations (95, 134), none of which are to be privileged (119) (also see Lyotard 1984:44, 81).

Rosenau suggests that this is a reaction to social scientists' "uncritical confidence in modern science and smugness about objective knowledge . . . (They) are uneasy with their more conventional colleague's (*sic*) uncritical acceptance of philosophical foundationalism, the Enlightenment heritage, and the methodological suppositions of modern science. . . . Those applying the post-modernism of the humanities to the social sciences . . . seek to . . . undermine their foundational claims" (9). Emphasizing the importance of the rejection of language as representational (see Baudrillard 1988:77-78, 80-92), "hostility to representation is linked to the post-modern preference for soft rather than positivist epistemology, to its aversion to objectivity, and to its insistence that difference is more important than sameness. Antirepresentation leads postmodernists to criticize modern scientific epistemology for its materialist conception of reality, for its separation of subject and object, and for its emphasis on generalization" (Rosenau 1992:96). In effect, there can be no valid or reliable mapping of social action or relationships by observing or interviewing actors.

Rosenau (1992) concludes that, if these arguments prevail, sociology "is no longer an attempt to understand and bring objectivity to a topic; rather, it is a vivid narration of a novel variety, *if it is anything at all*" (emphasis provided) (40). "Sociology . . . would be merely (a) literary endeavor, more comfortable in the arts than the social sciences" (172). Jonathan Turner (1992) has predicted that "if the programs of the critics were followed, sociology would be a mixed bag of rather pedestrian philosophizing, historical and empirical description, ideological debate and commentary, vague scheme-building, commentary on current (and past) events, extreme relativism, and a general doubt that we can know or do anything" (167). If so, sociologists are to be literary critics.

There are two aspects to the postmodernist argument. The first is a general rejection of orderliness or standardization in research. Implicitly it says that "in research anything goes." The second is a strongly anti-objective, anti-quantitative, and anti-empiricist theme. It rejects the kind of research that is common in American sociology — that typically is referred to as positivistic.[13] I consider the two in that order.

With respect to the first theme, I also have criticized standardized practices in sociological research and the assumptions that underlie it (Smith 1991a). Sociologists engage in vehement quarrels over what methods should and should not be used, insist on the virtues of their favored methods, and ignore the ontological and epistemological implications. I suggest that these conflicts usually involve largely meaningless, false dichotomies (e.g., quantitative and qualitative [see Hammersley 1992:159-73; Wolfe 1992:778], survey and comparative-historical research, micro and macro) and lead to poor research. I note that the very people that ethnomethodologists refuse to classify and generalize about, classify and generalize with ease in their own lives. Politicization and ideological quarreling on these issues leads to doctrinaire approaches to inquiry (also see Layder 1990:177-82; Wagner 1992:217).

Reflexivity and historical specificity may limit the validity of sociological propositions, but they do not invalidate application of the standard scientific model to social phenomena. If that were the case, it also could not be applied to physical and biological phenomena, many of which also are transient, historically specific, and change as they are observed. Entities, processes, and rules that cannot be perceived directly are inferred or assumed. Scientific inquiry requires abstractions that inform observations, not phenomena that are frequent, permanent, or observed directly. The claim that objectivity is impossible because the attribution of meaning requires prior assumptions (see Wortman 1987:171) is convincing only because it is expressed in a novel way (e.g., Baudrillard 1988:167-72). All science requires assumptions to focus observations and make them data. Assumptions are not problems; being unaware of them and unaware that they negate certainty are problems. Sociologists, however, flay one another for making assumptions — as if they could be avoided. Stephen Turner observes (1992:126) that "much of what passes for metatheoretical criticism in sociology . . . appeals to standards of . . . adequacy that would, if . . . fully formulated, be easily seen to be utopian." Unwarranted demands for perfection afflict and hamper sociology.

Because a moratorium on criticism and a return to the established model of scientific inquiry seem unlikely, I reemphasize sharing one's work as a universal norm in scholarly research. The responsibility requires scholars to have reasons for every decision made in a study. Research is inquiry marked by dilemmas (e.g, what is the question, where to look for an answer, what to observe for data). Investigators must decide how to resolve them and, if the goal of research is to achieve results that can be shared, investigators must have reasons that can be shared for every decision made. They must be pragmatic in accepting the compromises and assumptions that these decisions and their implementation require. However, they cannot be arbitrary or frivolous. Because the validity of conclusions rests on the decisions that determine the body of evidence and the rules for interpreting and evaluating it, rather than routinely relying on one

research model or procedure, methods must be created or chosen for clear, explicit reasons in each study (see Wagner 1992:206-207). Because conditions and questions change and methods must be designed for each study, no two studies could be exactly alike. Strict replication, except by imagination, is impossible. These prescriptions do not imply that there is only one world to be known through scientific inquiry (see Bauman 1992:29-31). That is an issue for empirical research subject to the same standards.

Any similarity of this position to the skeptical postmodernist position that "as far as method is concerned, 'anything goes'" (Rosenau, citing Heller and Feyerabend, 1992:117, see 123), is illusory.[14] The skeptical postmodernist view on research follows from the denial of the legitimacy of communal and supracommunal standards, and, therefore, of any obligation to adhere to them (See Bauman 1992:35). To accept the need to make decisions on unique, unexpected problems in research, despite appearances to the contrary, is not to accept the view that in research anything goes. Rather, it is to argue for research driven by moral imperatives. Researchers are morally obliged not only to choose standards for making decisions, to gather needed data to make them, and to decide, but also to reveal each decision and the reasons for it. Moreover, the empirical character of things must be assumed. If it is, they may be objectified and, thus, treated as finite, ultimately perceivable, and subject to observation and classification. To say that inquiry is legitimated only by being done in a community that subscribes to norms that require sharing one's work with peers, submitting it to their scrutiny for evaluation, and taking their judgment of it into account does not mean that the views of the majority or of elite scholars should prevail. Rather, it means there should be discourse among equals. Without such discourse, some could become authorities whose views could constrain knowledge (see Addelson 1988:126-29 on how the failure to treat philosophers as peers contributes to their authority). The position is not congenial to postmodernists. Lyotard (1984) considers such legitimation to rest on the "principle of consensus" (60), which he rejects as legitimation by paralogy. In his language we disagree on whether metanarratives have lost their force as a legitimating device (see Bohman 1991:126-32).[15]

With respect to the second theme, a critique of positivism in sociology, I would like to make two points. The first is brief and concerns the ambiguity of the term *positivism*. Giddens (1974) began his book on the subject by noting that "the term . . . has been used so broadly and vaguely as a weapon of critical attack, both in philosophy and sociology, that it has lost any claim to an accepted and standard meaning" (2). Eleven years later, Bryant (1985) concluded his study by saying that "variations in the characterisations of positivism . . . are sufficiently diverse to prompt the question whether it would not be better to expel the term from the vocabulary of sociology and other social sciences altogether" (184). Other scholars have not judged the situation as negatively, but have noted the variety of usages.[16] The ambiguity makes it difficult to address the criticism directly because each aspect of the conventional usage of the term — what Bryant (1985) and Ross (1991:390-470) refer to as instrumental positivism in American sociology — requires a somewhat different response. General attacks on positivism are vague and difficult to address succinctly.

The second point addresses the nature of differences between postmodernist and positivist positions. My impression is that the philosophical premises that underlie the critique are not new and may have been expressed earlier. Postmodernism may only be their most recent manifestation. Indeed, Rosenau's comment (1992) that postmodernists "abandon any attempt to represent the object of study 'as it really is' independent of the process of inquiry" (97) is remarkably like that in the textbook (Ramsperger 1942) that Hempel assigned my class in the philosophy of science and that informed us that "the positivist . . . is committed to the view that he alone exists (112)" and, therefore, that one cannot assume that other things exist. Hence, one can have confidence only in one's experience and "it will be better to speak of science as a description of experience" (103). Moreover, because scientific statements pertain to observations of specific kinds of objects, Ramsperger also writes that "'object' is here to be understood in a broad sense . . . Anything whatever that may turn up for an observer is subject matter for factual propositions" (109).

Despite basic differences on other presuppositions,[17] as a prescription for inquiry, these aspects of the philosophical positivism that influenced sociology have basic parallels with the postmodernist argument.[18] For postmodernists, texts may be anything (and there is nothing more than texts, which themselves may be illusory to others); for positivists objects may be anything (though a presumed real but directly unknowable something underlies what is observed). Both consider the individual and the individual's experience as the points of greatest confidence (though for positivists one is obligated to make the effort to share *in the event that others should prove to be real*, whereas for postmodernists others are no more than incidental[19]); both assert that all that can be known are the operations that provide observations. Skeptical postmodernists' denial of the possibility of true knowledge (Baudrillard 1988; Lyotard 1984; Rosenau 1992) resonates with positivists' insistence on uncertainty and error, though extremely skeptical postmodernists even deny the possibility of any truth.

The implications of these similarities and differences are instructive for assessing postmodernism as a critique of research. (1) Despite uncertainty, positivists expect inquiry to lead to assertions regarding whether evidence disputes an understanding of *an* unknowable, assumed reality and to an account of how the evidence was produced, a process that any trained investigator can evaluate and reproduce.[20] Postmodernists expect inquiry to produce a fictionalized narrative that satisfies its producer and others need not understand. (2) Postmodernists expect (Lyotard 1984:4) data transformation to electronic form to facilitate an undesirable exteriorization of knowledge. Positivism emphasizes the need to account, to provide, and to give directions. It prescribes the exteriorization that postmodernists distrust so that others can scrutinize and use one's work.

(3) Positivism and postmodernism also differ with respect to willingness to accept the existence of others, confidence in language to screen and represent experience, and openness to the possibility of shared understanding. Positivists are more willing, more confident, and more open than postmodernists. Skeptical postmodernists are more pessimistic and disillusioned.[21] Thus, whereas positivists assert that perceptions and the reality of the world may differ and whether they do is unknowable, postmodernists believe that the very idea of a

distinction is illusory or a delusion. (4) In regard to the experienced world, both positivism and postmodernism are political positions that argue for active rejection of the status quo. Both advocates and critics of postmodernism agree that postmodernists disagree on whether it merits a radical or conservative response (see Smart 1992:176-79). In contrast, positivism — ever since Saint Simon's and Comte's versions — has been optimistic and ameliorative, assuming that knowledge and understanding lead to improvement.

These similarities and differences reflect differences in the underlying assumptions of positivism and postmodernism, assumptions about reality and the relation of language to it. They lead to very different views of what we can know and act on. D'Amico (1992:142-47), a philosopher, also has noted similarities between positivism and postmodernism by analyzing changes in Carnap's position to show that postmodernist assumptions about reality and the role of language can be much more congenial to empirical sociology. If so, it emphasizes the importance of weighing the accuracy of the empirical claims about current social conditions that underlie the postmodernist argument before accepting its claimed implications for sociology and sociological research.

Some Hypotheses as to Why Some Sociologists Accept the Critique

This contrast of positions on communication and research suggests that the conclusions that skeptical postmodernists draw for sociology do not necessarily follow from the premises. Rosenau, Bauman, Smart, Crook, and Gurnah and Scott have all examined the critique and rejected it in total or in part for various reasons. However, their discussions largely are accommodating or sympathetic and neglect its most telling flaw — inconsistency. If skeptical postmodernists insist that it is impossible to establish certainty or organize coherently, and that dissensus, noncohesion, and nonintegration in the social world are inescapable, then why write? Who could read and understand them? How would they know? What would it matter anyway? Merelman (1991) despite deep concern that the critique challenges the meaningfulness of his research on culture and politics, suggests that criticism in this vein is "cheap philosophical gimmickry" (232). Inconsistencies between postmodernists' arguments and behavior, however, should not be ignored. Calhoun, for example, does not, when he charges (1992:268) that postmodernist criticisms of social scientists' assumptions on the nature of the reality are not presented in an arguable and defensible form. Because even flawed critiques can have pervasive effects on everyday life and social science practice, we must ask why the critique, despite its many deficiencies, maintains its appeal.

Perhaps postmodernism is just another passing scholarly fad. Nonetheless, it is important to understand its significance for sociology. One way to do this is to ask why it commands concern when phenomenological and Marxist critiques have been accommodated, if not always embraced. The concern probably is a reaction to postmodernism's nihilistic strain and extreme ramifications, qualities that are likely to deter its accommodation. Sociology is a discipline with boundaries; both discipline and boundaries are anathema to and targets of postmodernists. Given this, why do some American sociologists

accept and others tolerate as serious analysis an assessment by literary critics who preempt their empirical domain to make assertions that would be rejected if they were presented as based on sociological data? The situation makes little apparent sense.

No single or definitive factor can explain this forbearance. Many may be at work alone or in combination. Regardless of the fate of the postmodern critique, three are chosen here for consideration as hypotheses because of their implications.[22]

1. Criticism of sociological research as positivistic is correct.

2. Postmodernism derives credence from its European origins.

3. Postmodernism is considered philosophical analysis and sociologists defer to it as philosophy.

These hypotheses may not be the most important possibilities, are intuitively based, and are not independent. Positivism is a philosophy; postmodernism and positivism share key premises and progenitors; some philosophers are postmodernists; and some scholars interested in the relationship of philosophy and sociology consider philosophy superordinate and sociology subordinate. In reviewing reasons for entertaining these hypotheses and in the ensuing discussion I sometimes omit "some" before "sociologists" and "philosophers." It should always be assumed.

POSITIVISM

Giddens' (1974) comments on positivism include the phrase, "the term has become one of opprobrium" (2). Holmwood and Stewart (1991) wrote "that social theorists present positivism as a *discredited* methodology" (13) (emphasis theirs). Benton found it strategic to devote the first four chapters of his study (1977) of the relationship of philosophy to sociology "to a critical discussion of positivism, both as a philosophical theory and in its effects on sociological theory" (12). Critical theorists consider positivism to be the major source of sociology's weakness (Adorno et al. 1969) and Alexander urges sociologists to move on to a postpositivist mode (1992). Papers by Seidman, D'Amico, Stephen Turner, Brown, Calhoun, Aronowitz, and Alexander, among others, in Seidman and Wagner's volume either directly criticize positivism in sociology or indicate that others have subjected it to fundamental criticisms.

Criticisms of positivism primarily focus on its ontological and epistemological bases[23] or, in the case of instrumental positivism, the lack thereof. The disrepute of positivism is so widely accepted that some critics distance empiricism from positivism (though American sociologists tend to equate the two). Regardless of these critical nuances, positivism is a discredited position that still attracts bashers, postmodernists included. It is plausible that sensitivity to the identification of sociology with discredited positivism may incline some sociologists to tolerate the postmodernist critique, even if they would not accept it in whole or in part.

EUROPEAN ORIGINS

The centering of the postmodern critique in Europe also may be a factor in its acceptance by many European and some American sociologists.[24] Although American scholars (e.g., Frederic Jameson, Stanley Fish) have led in urging its assessment and application (primarily in literary and cultural studies), and an American, Daniel Bell, popularized the idea that a *possible* quantum shift in social life might signify entry into a phase beyond the modern, postmodernism's intellectual roots are in European philosophy, particularly modern French philosophy. Its targets, empirical sociology and adherence to the standard model of science, are primarily American. Few of the movement's analysts that I cite are American. Very few of the social scientists Rosenau cites as having accepted the critique and transformed their work are American. Nonetheless, sociology's European roots may incline American sociologists schooled in that version of the discipline's origins and/or otherwise favorably disposed toward European scholarship to accept postmodernism as a serious, legitimate critique. At the least, it might foster openmindedness toward postmodernism on the part of some who otherwise would have little interest in or knowledge of it.

PHILOSOPHY

I hypothesize that some American sociologists (1) consider postmodernism to be philosophy and (2) consider philosophy to be superordinate to sociology, and, consequently, (3) treat the critique with forbearance. The conclusion is speculative. The hypothesis assumes an attitude (perhaps with roots in the role of philosophy of science in our training in methodology) toward philosophy rather than an informed response to philosophy. It does not require all sociologists to know or assess the work of any American or European philosopher, only that many have that attitude. There are obvious objections to the hypothesis. The relevant evidence is mixed, at best. I review several aspects of the argument.

Why might philosophy even be relevant? Deconstruction is not unique to philosophy; postmodernist depictions of contemporary life are not philosophy; and its major spokespersons are in the humanities. Given this, why consider the postmodernist critique philosophy? Primarily because disciplinary identities often are ambiguous.[25] The ambiguity applies to postmodernists. Some say postmodernists are Sophists and that Sophists are antiphilosophy (D'Amico 1992:142; Doubt 1993:285). If Foucault, Lyotard, and Baudrillard[26] are Sophists and Sophists are anti-philosophy, then leading postmodernists may not be philosophers.[27] However, because postmodernism's intellectual roots are in philosophy, it is not clear what to attribute to philosophy and what to attribute to its use for ulterior ends. Where does philosophy end and its application and reinterpretation begin? Regardless of whether leading postmodernists are philosophers, many of the analysts cited here do consider leading postmodernists to be philosophers and all attribute their premises to philosophers (e.g., Nietzsche, Heidegger, Wittgenstein, Gadamer, Derrida). It is plausible, then, that many uninterested and/or unconcerned sociologists could take the critique to be philosophy.

Why assume a real or perceived status difference between the disciplines? The evidence is circumstantial. Philosophy is older and more firmly ensconced in the academy than sociology. From some perspectives it is a parent. There is a status difference in the projects of the two disciplines. Philosophy's project is to produce superior knowledge.[28] In contrast, sociology's project is more mundane and its critical style and subject matter more humble. For one thing, the norm of value-free inquiry has been interpreted as proscribing evaluations other than those of the appropriateness, validity, and weight of empirical evidence (though sociologists have valued the rationalization of social life as a goal[29]). For another, everyday life provides much of sociology's subject matter, and common sense explanations and understandings compete with its abstract theories and concepts for acceptance.[30] They may even enjoy equal status so long as the difference is not ignored and they are not confused (see Crook 1991:194-96; Layder 1990:137-69). There also is a major contrast in the status of conclusions. Philosophy's proofs are universal and certain; sociology's conclusions are relative and probabilistic (see Kilminster 1989). These conditions can promote inequality; inequality can breed deference; and deference can foster a subservient attitude.

The process and its effects can be illustrated by one of several perspectives on the development of American sociology. According to this reconstruction, there have been at least three instances of intensified relations with philosophy whose cumulative impact remains. The first peaked before World War II, when philosophy was among several influences on the subject matter of sociology. Philosophers seeking an ideal social organization and process reasoned from postulates about social and political conditions based on casual observations, assumptions, and speculation. Their arguments employed selected examples and needed systematic empirical grounding. For sociologists who accepted this as their project, it delineated sociology's substantive domain.[31] In a sense, for them, sociology served philosophy. The pattern was reflected in training. For example, Lichtenberger's *Development of Social Theory* (1936), a typical theory text, begins with Plato and Aristotle, and continues on to later Greeks and Romans, Aquinas, Bodin, Hobbes, Locke, Rousseau, Vico, and Montesquieu, arriving at Comte several hundred pages later. Lichtenberger referred to their work as social philosophy. It was prescriptive — offering models of how society might be organized and how it might operate — or analytic — postulating rules that govern social relations or social organization.[32] Though no longer pivotal for sociology's substance, Rawls, Habermas, Taylor, and other philosophers still raise central issues for the agendas of sociological specialties (e.g., criminology, mass communication, political sociology).

The second centered on positivism[33] and spanned the mid-1920s to the mid-1950s.[34] Positivism promoted deductive over inductive theorizing and prescribed formalizing any discipline that was to be a science. It set standards for concepts, propositions, and research. An SSRC monograph series on methods and the *Unified Encyclopedia of the Sciences* typified efforts to make social inquiry scientific. Positivism's impact endures in the dominance of experimental and survey methods, statistical analysis, and research patterned in a theory-grounded, hypothesis-testing mode. As compared with sociology as "empirical philosophy," the disciplines were more autonomous, but standards still came

from philosophy to sociology. The third is the current involvement with postmodernism.

In each case, an autonomous philosophy has provided agendas, standards, and/or prescriptions for sociology.[35] In each, philosophy has had almost hegemonic status for sociology. Bourdieu's experience of a loss of professional status and self-esteem in switching from philosophy to sociology (Wacquant 1992:45) dramatizes how such inequality can he detrimental for any sociologist. In contrast to sociologists' sensitivity, many philosophers have been indifferent to, if not disdainful of, sociology.[36]

Social philosophy's role may have waned,[37] but philosophy's superordinate status was and is manifested in various ways.[38] Durkheim (1933:370-71) designated philosophy as collective conscience to the sciences even though he deplored its abstractions (Ragin and Zaret 1983:732). Adorno (cited by Crook 1991), with experience in both disciplines, wrote that philosophy "take(s) its specific scientific material preponderantly from sociology . . . (because) . . . the idea of science is research . . . that of philosophy is interpretation" (102). Interpretation is the source of meaning and significance. Merleau-Ponty ([1960] 1974:95), although urging an end to the segregation of the two disciplines, advised sociology to look to philosophy for understanding (107). Kilminster (1989) charges that philosophy is losing its role, and, to save itself, imposes a hierarchical division of labor on the social sciences. All the above points lend plausibility to the hypothesis that some sociologists might consider the critique philosophy, might consider philosophy superior, and, thus, might tolerate, if not accept, the critique.

How May the Situation Make a Difference to Sociology?

The factors specified in the hypotheses — sensitivity to being considered positivists, acceptance of fundamentally flawed arguments if they have European origins, and deference to philosophy — all portray sociology as a subordinate, weak discipline. Because subordination and weakness are pernicious, I consider their implications for the discipline and its future.

There are two related implications of an inequality between philosophy and sociology for sociologists' forbearance toward the critique and its impact on the discipline. (1) It inclines sociologists to accept passively and not scrutinize critical philosophical analyses. (2) Therefore, most sociologists who consider postmodernism to be philosophy tolerate it. The first could be a by-product of longstanding patterned relationships that plant the seeds of inequality.[39] Philosophers propound standards and rules for the structure of theory, the relation of theory and research, and the conduct of empirical inquiry that sociologists accept. Philosophers' critiques of sociology can be uninhibited and invidious (see Bohman for an exception); sociologists accept them. Few reciprocate.[40] In essence, philosophers' role is to decree and criticize, sociologists' role is to accept. The relationship may explain why there is no sociology of philosophy.[41] A sociology of philosophy would reveal that philosophers' work is no purer than sociologists' and that philosophers also do empirical

research[42] that, by the standards of evaluation for empirical research, is no better than sociologists'.

The practice of philosophy goes unexamined. The discipline's "ivory tower" reputation — socially and historically unsituated, pursued in isolation from the real world — diverts scrutiny. A sociology of philosophy would examine social influences on philosophers' work, as Gouldner did in *Enter Plato* (1966), or provide socially situated interpretations of that work, like those of Zeitlin (1968) and Coser (1977). It would show that philosophers, like other scholars, respond to and try to influence their social, economic, and political environments.[43] The ivory tower myth shields philosophy from the sociology of knowledge and discourages a sociology of philosophy. This helps sustain the inequality.

Ontology and epistemology draw attention to the assumptions and the irreducible uncertainty in empirical research.[44] They become ammunition that sociologists use to castigate one another for striving to be scientific. Sociologists' critiques of research practice tend to be more extreme for sociology than for other disciplines. Morbid, destructive self-criticism on the basis of utopian philosophical standards pervades the discipline.[45] Charges of scientism grow in force and volume with the continued discrediting of positivism (see Peters and Rothenbuhler 1989). Studies of scientists' actual work patterns are cited to legitimate alternative approaches (see studies cited in note 6). Sociologists become chronically uneasy with reminders that they never can be sure that anything they presume to know outside themselves exists. Concern that disciplinary concepts lack meaning or currency in everyday life fuels nagging doubts that their referents exist. These debilitating insecurities are magnified by the postmodernist charge that the inability to verify meanings vitiates their work. In this way, the confluence of a sense of inferiority to philosophy, the equating of postmodernism and philosophy, and insecurity in the face of elegant European theorizing and criticism not only encourages tolerance of the critique but also undermines confidence in sociology.

Postmodernists' claims that sociology is a political project also disturb and divert sociologists even though postmodernism, too, is a political project (as well as an application of philosophy). They lose sight of the fact that the charge of being political is a truism that applies to and does not invalidate any discipline. Every field of knowledge, including philosophy, influences and is influenced by its circumstances. Sophisticated scholars know this. Even if such concepts as society and culture were invented to serve powerful social and political interests,[46] their survival rests on how well they predict and explain and on their relation to other explanatory systems, not on their origins.

The crux of the matter is that postmodernist contentions about whether there should or can be social sciences as now practiced rest on empirical claims and interpretations of their meaning. The qualitative shift that postmodernists claim has occurred in people's relation to events and to one another is empirical and rests on anecdotal rather than systematic evidence. Its meaning is postulated. Baudrillard's exposition (1988:141-48) of a transformation of the real to fantasy and the hyperreal is brilliant polemic. The conception of a postmodern world need not be wrong because it is based on selected examples for data and on *ad hoc* interpretations for meaning, but it must be tested with systematic data and other frameworks for interpretation. No matter how insightful, a critique

built on empirical claims is subject to evaluation on the basis of systematic empirical data. Entertaining an assessment from an intellectual conceit that passes as philosophy and is particularly inimical to sociology (Smart 1992:180-81) perpetuates sociology's inferior status and weakness. It also is an unwitting surrender of turf.

The aggressive colonizing of the sociologies of mass media and culture by postmodernist cultural and literary scholars is only the latest incursion invited by sociology's institutional vulnerability.[47] The sociologies of organization, crime, family, politics, and public opinion already have been or are being appropriated (see Smith 1991a:2). There may be several reasons for this situation.

1. It may be inevitable for a discipline dominated by positivism; overconcern with methods can lead to neglect of substance.[48]

2. It may be an indirect result of the view that sociology is overarching and generalizing and, therefore, has expertise over all social reality. If so, sociologists are victims of their own smug complacency.

3. Sociology may have grown too big, and, because there is more knowledge than can be digested, it has fragmented in substance and purpose. This facilitates raids because most sociologists know little and care less about what is happening in distant specialties (Collins 1992:187-88).

Regardless of the reason, when resources are scarce, being correct about the social science hierarchy or the proper way to do sociology is irrelevant. Scholars who worry that their discipline may have nothing new to say but still must compete for students and tight budgets, seek and pursue any opportunity aggressively. All of sociology is appealing turf to invade.[49] For a discipline whose institutional support is primarily academic, loss of turf portends a bleak future.

Conclusion

Even if postmodernist premises are suspect, the possible prevalence of the conditions that the analysis presumes cannot be ignored. People increasingly may be divorced from reality, assimilation may not reduce heterogeneity, diversity may be increasing, integration may be an unattainable goal, and change and dissensus may describe social reality better than stability and consensus. If so, these social conditions will influence sociological theories, models, and concepts (see *Sociological Theory* 1991, 1992). Giddens's concept of structuration, Bourdieu's method of epistemic reflexivity, and Touraine's vision of society as social movements exemplify the possibilities. Computers can expedite these developments.[50] Research also will change. Methods increasingly will be shaped and judged in terms of their intrinsic relation to subject rather than their conformity to abstract method-sets (e.g., survey, experimental, comparative-historical method). The reasons for choices rather than specific methods and techniques will be taught; data gathering will be conceptualized as social behavior rather than as inflexible rules of action.

Even if the dire implications of the postmodern critique for sociology are neither logically nor empirically necessary, we may not be assured of escaping its influence. Jonathan Turner (1992), after assessing the critique, concluded that:

> with the invasion of so many new intellectual stances from Europe, as well as the fostering of those initiated in the United States, theoretical activity is partitioned into too many "camps," "orientations," "perspectives," or "paradigms" for any general theory, cutting across this diversity, to ever be possible. Such is especially likely to be the case in a field that will not even take itself seriously as a science, preferring instead to be a mixture of history, philosophy, social commentary, ideological advocacy, and empirical description. (175-77)

Collins (1992:187-88), emphasizing fragmentation, also is pessimistic. I have argued that the discipline stagnates while sociologists fight over ontological and epistemological matters (also see Wagner 1992:217).[51] It is a mistake to focus only on postmodernist criticism when dissension is rife.

The postmodernist influence on sociology could lose force more quickly than positivism did. However, that is secondary. Defusing the critique will not make sociologists less prone to accept unwarranted criticism. In concluding their analysis, Gurnah and Scott (1992) write that "the entire modernity versus postmodernity debate has done little to change the relationship between sociology and philosophy. The old subordination remains" (161). Though neither this analysis nor theirs might have been necessary if the disciplinary relationship had been more equal, they are forceful reminders to sociologists to accept the dynamism and complexity of their subject — and think for themselves.

Notes

1. Nonetheless, it was the topic of a symposium in *Sociological Theory* (1991) to which Seidman, Alexander, Antonio, Lemert, and Richardson contributed. Several other articles on the subject also have appeared in that journal. Both Aronowitz (1988) and Gitlin (1989) have written on postmodernism and politics. Seidman and Wagner's edited volume (1992) includes papers by three of the above authors as well as six other sociologists. When people as respected as Collins, Gitlin, Calhoun, and Alexander address the critique, it seems reasonable to treat it seriously even if it is not widely accepted.

2. See Lemert's (1992) impassioned plea for a sociology that does not employ universal categories because that would imply a realist epistemology when, in fact, real people "make very general statements . . ." despite "the fact that they do not and cannot believe in a world that is whole or real outside of what they experience and express" (40-41).

3. An issue of *Theory, Culture and Society* devoted to postmodernism is replete with discussions of these differences (see Bauman 1988:217; Featherstone 1988:197; Kellner 1988:241). Bernstein (1992) says of the situation that "the terms 'modern' and 'postmodern' are not only vague, ambiguous, and slippery, they have been used in conflicting and even contradictory ways" (200). See Smart (1993:11-39) for a full discussion of the many usages of the term.

4. Baudrillard (1988:57-97) considers the alienation and fractionation of value and its divorce from that which is valued under capitalism as crucial. Lyotard (1984) identifies technological innovations that alter and depersonalize that which is valued as being among the crucial factors expediting the transformation. Both claim that computing has enhanced the commodification of knowledge and that the broad influence of that process is a major feature of the postmodern world.

5. Rosenau distinguishes skeptical and affirmative postmodernists and moderates and extremes of both types, and notes that the high correlation between (1) complete rejection of or revision of modernism and (2) extreme or temperate expression makes for very varied characterizations of postmodernist commentary (8, fn. 11). Smart (1992:167), for example, prefers to distinguish European and American types. Because their position is the most extreme and carries the most drastic implications for sociology, I focus on versions that Rosenau calls skeptical.

6. The critique's relevance to sociology may reflect its Nietzschean roots. Lepenies (1988) refers to Nietzsche as "the anti-sociologist. . . ," (41) who "saw in Comte . . . a continuation of . . . eighteenth century . . . 'sensualism in epistemology'" (40-41). Its acceptance by sociologists as a legitimate critique may have been expedited by its use of deconstruction. Deconstruction complements the social constructionism that appeals to many sociologists. It also is employed by Marxism and feminism, both of which argue that conventional sociological depictions of society are constructions that serve the interests of elite white males. Accommodation to these critiques may have facilitated postmodernism's acceptance (see Nicholson 1992; Richardson 1993).

Relativism, an influence on sociology even earlier than Marxist and feminist criticism, also may contribute to some sociologists' forbearance with postmodernist claims. Relativism, when applied to the sociology of knowledge, has contributed to what has come to be known as the sociology of scientific knowledge, which treats the dominant conventional representation of scientific practice as a fictive construction that does not represent actual scientific practice (see, e.g., Barnes 1975:ch. 5; 1982:ch. 2; 1985; Lynch 1985; Latour & Woolgar 1979). The more general concern with the discontinuity between the actual and claimed practices on which assertions are based also was a factor in the earlier development of ethnomethodology (see Barnes 1982:94-101; Lynch 1985:6-10; S. Turner 1994:38-43). Practitioners of a discipline that argues that social factors shape and impinge on behavior, beliefs, and values are, perhaps, more open than practitioners of other disciplines to entertain the possible merits of, if not accept, the postmodernist critique. See Stephen Turner 1994 for an explication and discussion of the deeper roots of these connections.

7. Cf. fn. 26. Charles Lemert, in a summary of the postmodernist argument consistent with all the versions and aspects I have examined, states (1992:23-24) its principles as follows:

1. Reality is discussable — not much more than this.
2. Language is primary.
3. Reality is figurative, available only in language.
4. In language one can say any and everything, including the most general of things.
5. One can never escape one's language, including all other general statements about things.
6. All the above is a historical event more than an ideological position or a logical argument.
7. Hence, postmodernism is both the end of modernity and dependent upon modernity for its language.

8. Rosenau (1992:95) and Bauman (1992:33) cite Baudrillard as claiming this.

9. Rosenau (1992) says "The skeptics are dubious about the possibility of communicating altogether" (18).

10. Bauman (1992:54) refers to socialization and culture (including language) as derandomizing the voluntary actions of free agents.

11. Baudrillard (1988) would reject this definition. Though we agree that the media entail a one way use of power to influence and, therefore, are not engaged in communication, he says that communication may have utility only "if one defines it as *anything other* (emphasis mine) than the simple emission/reception of information" (207). This is the antithesis of our definition.

12. Zaret reports (1980:85; 1985:19-40, 53-60) that as early as the late 16th century, the nascent English popular religious press was important in forming the religious concepts of Puritan laymen by mediating their relations with clergymen.

13. Using Rosenau's (1992) index, I easily found at least a half-dozen references to positivistic social science as a specific target of postmodernist criticism.

14. Gellner's (1992) understanding of Feyerabend is more nuanced and sympathetic. "It isn't so much that 'anything goes,' as that it becomes difficult to explain why this should *not* be so" (110-11). Feyerabend was a student of Popper. This makes credible D'Amico's (1992:145)

surmise that Feyerabend's methodological anarchism actually derives from Carnap. If so, Lyotard (1984:76) may not use the phrase as Heller and Feyerabend do when he refers to postmodern cultural practices as being guided by the idea that "anything goes." See Morris' (1992:259) criticism of David Harvey for misunderstanding Lyotard's usage.

15. If affirmative postmodernism prescribes "feelings, personal experience, empathy, emotion, intuition, subjective judgment, imagination, as well as diverse forms of creativity and play . . . as alternatives to scientific method" and leaves the "content of these terms and their methodological significance . . . relatively vague and difficult to communicate to others" (Rosenau 1992:117), then it also differs from the accountability prescribed by radical social theory (see Crook 1991:200-14). That is a responsibility to be aware of and reveal one's interests and values and how they may affect decisions on topics for inquiry, observations, and conclusions. Although not always honored (see Crook 1991:201), conventional scientific norms also permit accountability. Skeptics deny there are alternatives.

16. Cf. Benton 1977; Brown 1990; Holmwood & Stewart 1991:11-25. Bryant (1985:2) credits Percy Cohen with a more systematic analysis of its meaning. Most critics, including postmodernists, use positivist as an epithet to refer to empiricism, quantification, the use of statistical models, formalism, and overconcern with procedures, singly or in combination. Because they identify these features of social science with survey research and statistical analysis, "positivist" often is used to describe the latter. All these referents are subsumed by instrumental positivism (Bryant 1985:133-73).

17. The meaning of similarities in the face of different underlying assumptions is debatable. However, postmodernists' efforts to address a public indicate that, regardless of rhetoric, they seek understanding and support from others. If so, how they arrive at empirical assertions becomes crucial and both parallels and differences in positions may be significant.

18. Wittgenstein has been identified as a progenitor of both. He sympathized with the program of the Vienna Circle (Bryant 1985:110) and is cited by postmodernists as a source for their arguments.

19. This may reflect the contrast between the postmodernist position that we cannot overcome the uncertainty and unreliability of *any* attempt to communicate and the positivist belief that they can be overcome by clarity, precision, and order. These positions reflect a deeper division on materialism.

20. Ramsperger writes that "Science is at least an orderly account of experiences; this implies that it must select and classify experiences in such a way as to *give* them order (103) . . . The purpose of science . . . is to facilitate the prediction and control of experience by providing propositions and formulae which express, in the most economical manner, the relationships between observable objects and events (106). . . . The cognitive meaning of a statement consists, then, in the directions it gives for getting perceptions of a definite sort (107) . . . No statement about existence is meaningful unless it contains, at least implicitly, directions for the observation of a specific kind of object" (109).

21. Calhoun (1992:265) has described many as happy relativists who eschew the gloom it can bring. However, pessimism may be too mild a term for Baudrillard (1988), who has written that "Nothing can insure us against fatality, much less provide us with a strategy" (203).

22. Those aspects of the sociological tradition that have encouraged the constructionist view of knowledge and current developments in the sociology of scientific knowledge in particular (see note 6) are not discussed here. They function more to provide a general climate in which the factors selected for discussion may take on increased relevance.

23. On these grounds, Brown (1990:322) has described positivistic absolutism as self-refuting.

24. The conjunction of European criticism and American empirical social science may be more than coincidental. The two long have been in a state of tension arising from matters that have little to do with issues of validity. The feelings of at least some American scholars are exemplified by Stinchcombe's accusation (1993) that British sociologists are parochial, and by Jonathan Turner's (1992:156) claim that it is unfashionable to identify as a positivist because American theory circles are inundated with European sophistication in the form of phenomenology, hermeneutics, structuralism, and critical theory. The tension on the European side may reflect concern about inferior intellectual resources, influence, and power that have resulted in

the shift around 1930 of major advances in social science from European to American sources (Deutsch, Markovits & Platt 1986).

25. Addelson (1989) refers to Peter Winch as a philosopher who already has understood and implemented, even if insufficiently, her call for philosophers to become sociologists. She says of her own semiconversion that "I was trained as an analytic philosopher, and I started to become a sociologist in 1974" (140). Gellner (1973) ascribes to Winch the view that "philosophy and sociology (in a broad sense) are very similar, if not almost identical" (48) and, given Winch's reasons for that view, that "the implicit corollar(y) . . . is . . . that philosophy really is applied sociology." Gellner himself has held academic appointments in philosophy, anthropology, and sociology. Others (e.g., Schutz, Foucault, Habermas) have had similarly fluid careers. Several sociologists (Giddens 1984:xvii; Lemert 1992:18-19; Doubt 1993:285) say that social theory inevitably involves analyses that spill over into philosophy and that much sociological theorizing becomes indistinguishable from philosophizing.

26. Nonetheless, Bogard (1990:1), for one, refers to Baudrillard as a social philosopher.

27. Calhoun (1992:265-76), Stephen Turner (1992:102), and others, identify the rejection of philosophical foundationalism and the depiction of most sociological theories as foundational to be essential characteristics of postmodernism. D'Amico (1992:142) likens most philosophy to foundationalism. Postmodernists reject foundationalism because foundational claims are political commitments to idealistic positions. Assertions as to the acceptability and priority of foundational claims, in turn, are only exercises of power. Calhoun maintains that alternatives to foundationalism just don't appear. In any case, in this light the postmodernist position is antiphilosophy as well as antisociology.

28. Bauman (1992:115) interprets the *modern* and *postmodern* modes in sociology and philosophy as legislative and interpretive. He cites Plato (quoting Socrates), Kant, and Descartes to support the argument that philosophers see their role as pursuing inquiries that rise above everyday life so as to derive *superior* knowledge that can be used to legislate in the interests of the people (116-17). Ruth Macklin (1985:276), a philosopher, classifies philosophical writings as being of five types, at least the first three of which — defining terms and analyzing concepts; attending to the logic of arguments, detecting fallacies, and uncovering assumptions; and analyzing and interpreting writings from both within and without philosophy — imply that philosophy has methods and capabilities that justify analyzing, criticizing, and prescribing for other disciplines.

29. The development of sociology can be construed as one effort to achieve the Enlightenment's goal of facilitating a more rational social life (see Smart's discussion of Habermas's view of modernity, 1992:162). Bauman (1992:39) describes Giddens' assessment of sociology as being constituted by the "rational analysis of society (which itself is) . . . subject to the processes of continuing rationalization." The same aspiration is expressed in James Coleman's recent (1993) presidential address to the American Sociological Association. Rickman (1983), a philosopher, insists that "any systematic study of society should be grounded by reason . . . [and] philosophically treated" (3).

30. Bauman (1992) says sociology has inherited philosophy's project of producing superior knowledge on which to base legislation of a better life. It has pursued the project by "the critique of common sense" and by proposing ways to avoid social "dis-order." Eventually, this has "led inexorably to the denial of the human capacity for generating adequate self-knowledge" (121-22). He claims that this was both Durkheim's and Weber's intention and is reflected in their explanations for their methodologies. He asserts that for orthodox sociology "lay knowledge of society members cannot be trusted as representation of truth. To put it bluntly, people on the whole do not know what they are doing or why they are doing it" (124). Earlier, Simmel had noted the co-existence and irreducibility of the syntheses of social actors and of analysts. Neither is privileged because each is conditioned by the interests of the person in the role (see Crook 1991:193). The distinction between actors' and analysts' perspectives is crucial in ethnomethodology and phenomenological sociology. Compare also Misgeld's analysis (1986:83-84) of why natural language and "common sense reasoning could not be trusted to give (sociology) a correct view of society" (84).

31. Addelson (1992:141, fn. 4), for example, writes with respect to sociological symbolic interactionism that it has "philosophical roots in American pragmatism of the early twentieth century . . . We must follow the pragmatist philosophical position forward to one of its natural outcomes in a sociological tradition of today . . . Becoming a sociologist requires working

within a research tradition that has corrected and extended the original, philosophical ideas of its founders by making them empirical."

32. The philosopher William Montague (1927) elaborated the concept of social philosophy. If philosophy is "the attempt to give answers to the kinds of questions that have not yet been answered with definiteness and certainty by science," social philosophy attempts to answer two main questions: what is society, and what goal or end should be used to evaluate social practices. As for the relationship, "philosophers are the advance guard for the army of scientists. They spy out the new land as a necessary preliminary to its successful occupation" (466).

33. Positivism reflects the analytic-synthetic distinction of the Enlightenment and is manifested in logic, formal languages, and meta-mathematics. It is applied to science as conceptual and linguistic analysis. See Addelson (1992:123).

34. Throughout this article I conflate European logical positivism in philosophy and American instrumental positivism in sociology. The first is a response to a variety of traditional problems in European philosophy (Bryant 1985:109-11). The second (Bryant 1985; Ross 1991:390-470), a somewhat earlier self-generated position, has roots in an optimistic commitment to what Ross calls scientism as a means to social betterment. It may have derived from John Dewey and his school of philosophy. The confusion may owe to the fact that many European positivists had come to the United States during the 1930s and held university appointments in the same institutions as sociologists who were instrumental positivists. When I was an undergraduate in sociology in the early 1940s, the two were not being clearly distinguished and logical positivism was being cited as a justification for research practice.

35. The pattern also has been noted by Bauman and by C.G.A. Bryant. Bauman (1992) writes that "the declaration of intent associated with philosophy impelled and agitated by legislative reason was, overtly or implicitly, an anti-sociological manifesto . . . Philosophy . . . left sociology a choice between the role of handmaiden . . . or facing the prospect of dishonourable discharge without references" (143). Bauman goes on to say that interpretive philosophers have drastically changed this situation, but expresses it as their decision. This still implies hierarchical relations. Given the pervasive long-term influence of established institutions, the idea that hierarchy is easily changed on the basis of purely intellectual decisions is questionable. Bryant (1992) writes that "the relationship between sociology and philosophy has long been contested by sociologists and philosophers . . . [S]ociology can do without philosophers but not without philosophy. Sociology inescapably raises issues of ontology, epistemology, and normative theory. Either sociologists must rely implicitly or explicitly on positions articulated by philosophers, or they must engage the issues themselves and come up with something better" (137). His discussion is based on what he describes as Kilminster's brilliant analysis of how philosophers have used their greater status to impose their definition of the relationship.

There are, of course, exceptions. Empiricists like William F. Ogburn ignored philosophy. Many historical sociologists, who would disidentify with Ogburn, also emphasize detail and particularity in their work and ignore or contest the relevance of philosophy for sociology (see Skocpol 1987).

36. Cf. Gellner's (1973) statement that the philosopher Peter Winch:

> checks his conclusions by arguments . . . with the general methodological and programmatic views of major figures in the thought *about* the social sciences. But the social sciences are also a *reality*, and one which, however imperfect, did not wait for recent philosophic revelations for permission to actualize itself. To this reality Mr. Winch in his rather Platonic procedure does not pay much attention. Had he done so, he . . . would at the very least have avoided the somewhat awe-inspiring posture of a philosophic Superman flying in from outside and setting things right for the earthbound social scientists. (47-48)

Addelson, D'Amico, Bohman, and Merleau-Ponty, in contrast, respect sociology and consider it in some respects an equal or even a leader in a common enterprise.

37. Amitai Etzioni, Philip Selznick, Herbert Gans, Jeffrey Goldfarb, and Robert Bellah are all American sociologists who argue for sociological analysis that is normative and can play a policy role. They might be called social philosophers. The nature of social philosophy is captured by Lemert's (1992:33-34) distinction between (epistemologically grounded) sociological theory and (politically grounded) social theory. It is a permeable distinction that reflects the

ambiguity of the line between philosophy and sociology. If Selznick's work on community (1992) had been done as a service to Taylor's analyses of modern life (1989, 1990, 1991), it would provide an example of contemporary social philosophy.

38. Lyotard (1984) makes this point more subtly. "To the extent that science . . . seeks the truth, it is obliged to legitimate the rules of its own game. It then produces a discourse of legitimation with respect to its own status, a discourse called philosophy" (xxiii).

39. A sociological inferiority complex toward philosophy might have roots in sociology's failure to become established in European institutions at the height of its classical period in the early 1900s (Wagner 1990). Wagner cites Croce's 1906 description of sociology "'as a barbaric positivistic incursion into the domain of philosophy and history'" (225), and asserts that, as a newcomer, sociology would not be welcomed by such established disciplines as philosophy and that "early sociologists were aware of this institutional problematique" (232-33). Some philosophers acknowledge that there is an unbalanced relationship between philosophy and the social sciences and propose restructuring it as a relation between partners (see Bohman 1991:235).

40. Gurnah and Scott (1992) provide a rare exception in saying that "the continued interlacing of philosophy and sociology distorts sociology and limits its critical impact . . . The tension between sociology as an empirical investigation of the culturally and historically specific and philosophy as a reflection of the constant and universal has haunted sociology from the start . . . Our thesis is anti-philosophical . . . [S]ocial research is limited if it is structured by philosophical concerns" (1). Our reasons for judging the relationship negatively differ, however. Gurnah and Scott want to maximize sociology's critical force and see any association with or grounding in positivism as foundationalism, which, they argue, makes sociology vulnerable to flaws that come from an association with philosophy. I (1991a) argue, in contrast, that a sociology influenced by positivism also is critical. Postmodernists (or others loath to assume an empirical reality) differ from both positivists and critical theorists, claiming that any or no assumption about reality may be made and, in any case, that it does not matter.

41. Bauman's (1992) discussion of the relation of sociology to philosophy refers to social philosophy but not to a sociology of philosophy.

42. Before philosophy and psychology split over empirical research, G.H. Mead grounded his graduate work in philosophy in empirical studies of the mind. Later, members of the Frankfurt School did empirical studies (cf. Aronowitz 1992:291-94) of the media (Adorno, Lowenthal) and prejudice (Adorno, Horkheimer); Kaplan, Rudner, and others did content analyses (applied hermeneutics) and experiments (applied philosophy of science and knowledge).

43. Shalin (1992:239-44) shows how an interest in diagnosing and controlling political developments shaped the work of continental philosophers from the French Revolution to the rise of the Nazis.

44. Jonathan Turner (1992:156-57) attributes problems in sociological theorizing to the detrimental effects of philosophy on the work of sociologists. Stephen Turner (1992) places its source in the rigid application to sociological research and theorizing of absolute standards that derive from philosophy.

45. Compare Stephen Turner's comment quoted in this article in the methodology section.

46. For example, see Smith (1991b) on the promotion of a national culture by political and cultural elites as a project to consolidate power.

47. Peter Wagner's analysis of sociology's failure to become established in European scholarly institutions employs the argument that disciplines legitimized themselves by claiming a sphere of interest not already claimed by an established discipline. Post-World War I German sociology was an exception. Von Wiese succeeded by arguing "for a reductionist conception of sociology that would avoid all possible conflicts with other disciplines" (236). Wagner notes that only in the United States, where disciplinary boundary setting and professionalization were even more extreme, did that concept play a similar role in establishing sociology institutionally. Subsequent developments varied by country, so each must be understood in terms of its own conditions — structure of the university system, role of the state, etc.

48. However, Kaplan (1964) espouses the view that attention to methods requires attention to substance.

49. David Zaret has observed that *tour de force* literary critiques based on textual analysis are a low cost tactic for invading turf (Personal correspondence).

50. Postmodernists decry what they believe to be the impact of computers on knowledge and thought. It is intriguing to speculate on what postmodernism would be like if analog rather than digital computers had dominated.

51. Two critics of conventional empirical research (Bryant 1992a, 1992b; Fuller 1992,1993) are currently embroiled in a bitter debate in which each claims that his version of reflexive sociology is the only appropriate model for sociological inquiry. Ironically, to support his case, each cites Duncan, Blalock, and Lieberson, all of whom urge care and creativity in reasoned methodological critiques of specific weaknesses in quantitative research practice.

References

Addelson, Katherine P. 1990. "Why Philosophers Should Become Sociologists (and Vice Versa)." Pp. 119-47 in *Symbolic Interaction and Cultural Studies*, edited by Howard S. Becker and Michal M. McCall. University of Chicago Press.

Adorno, Theodor W. 1976. *The Positivist Dispute in German Sociology*. Harper.

Alexander, Jeffrey. 1992. "General Theory in the Postpositivist Mode: The 'Epistemological Dilemma' and the Search for the Present." Pp. 322-68 in *Postmodernism and Social Theory: The Debate over General Theory*, edited by Steven Seidman and David G. Wagner. Blackwell.

Aronowitz, Stanley. 1988. "Postmodernism and Politics." Pp. 46-62 in *Universal Abandon? The Politics of Postmodernism*, edited by Andrew Ross for *Social Text*. University of Minnesota Press.

____. 1992. "The Tensions of Critical Theory: Is Negative Dialectics All There Is?" Pp. 289-321 in *Postmodernism and Social Theory: The Debate over General Theory*, edited by Steven Seidman and David G. Wagner. Blackwell.

Barnes, Barry. 1974. *Scientific Knowledge and Sociological Theory*. Routledge & Kegan Paul.

____. 1982. *T.S. Kuhn and Social Science*. Columbia University Press.

____. 1985. *About Science*. Basil Blackwell.

Bauman, Zygmunt. 1988. "Is There a Postmodern Sociology?" *Theory, Culture and Society* 5:217-37.

____. 1992. *Intimations of Postmodernity*. Routledge.

Baudrillard, Jean. 1983. *In the Shadows of the Silent Majorities*. Semiotext(e).

____. 1988. *Jean Baudrillard: Selected Writings*, edited with an Introduction by Mark Poster. Stanford University Press.

Benton, Ted. 1977. *Philosophical Foundations of the Three Sociologies*. Routledge & Kegan Paul.

Bernstein, Richard J. 1992 *The New Constellation: The Ethical-Political Horizons of Modernity*. MIT Press.

Bogard, William. 1990. "Closing Down the Social: Baudrillard's Challenge to Contemporary Sociology" *Sociological Theory* 8:1-15.

____. 1992. "Postmodernism One Last Time: A Comment on Seidman et al." *Sociological Theory* 10:241-43.

Bohman, James. 1991. *New Philosophy of Social Science*. Polity.

Brown, Richard Harvey. 1990. "Symbolic Realism and the Dualism of the Human Sciences: A Rhetorical Reformulation of the Debate between Positivism and Romanticism." Pp. 320-40 in *The Rhetorical Turn: Invention and Persuasion in the Conduct of Inquiry*, edited by Herbert W. Simons. University of Chicago Press.

Bryant, Christopher G.A. 1985. *Positivism in Social Theory and Research*. St. Martin's Press.

____. 1992. "Positivism without Philosophy? The Case of Giddens' Structuration Theory," *Sociological Theory* 10:137-49.

Bryant, Joseph M. 1992a. "Positivism *Redivivus*? A Critique of Recent Uncritical Proposals for Reforming Sociological Theory (and Related Foibles)." *Canadian Journal of Sociology* 17:29-53.

____. 1992b. "Towards a Respectable, Reflexive, Scientific Sociology: A Note on the Reformation Required." *Canadian Journal of Sociology* 17:322-31.

Calhoun, Craig. 1992. "Culture, History, and the Problem of Specificity in Social Theory." Pp. 144-88 in *Postmodernism and Social Theory: The Debate over General Theory*, edited by Steven Seidman and David G. Wagner. Blackwell.

Coleman, James S. 1993. "The Rational Reconstruction of Society." *American Sociological Review* 58:1-15.

Collins, Randall. 1992. "The Confusion of the Modes of Sociology." Pp. 179-98 in *Postmodernism and Social Theory: The Debate over General Theory*, edited by Steven Seidman and David G. Wagner. Blackwell.

Coser, Lewis A. 1977. *Masters of Sociological Thought*. 2d ed. Harcourt Brace Jovanovich.

Crook, Stephen. 1991. *Modernist Radicalism and Its Aftermath: Foundationalism and Anti-Foundationalism in Radical Social Theory*. Routledge.

D'Amico, Robert. 1992. "Defending Social Science Against the Postmodern Doubt." Pp. 137-55 in *Postmodernism and Social Theory: The Debate over General Theory*, edited by Steven Seidman and David G. Wagner. Blackwell.

Deutsch, Karl W., A.S. Markovits, and J. Platt. 1986. *Advances in the Social Sciences: What, Who, Where, How?* University Press of America.

Doubt, Keith. 1993. Review of *Postmodernism and Social Theory: The Debate over General Theory*, edited by Steven Seidman and David G. Wagner, and *Intimations of Postmodernity* by Zygmunt Bauman. *Contemporary Sociology* 22:285-86.

Durkheim, Emile. 1933. *The Division of Labor in Society*, translated by George Simpson. Macmillan.

Featherstone, Mike. 1988. "In Pursuit of the Postmodern: An Introduction." *Theory, Culture and Society* 5:195-215.

Fuller, Steve. 1992. "A Plague on Both Your Houses: Beyond Recidivism in the Sociological Theory Debates." *Canadian Journal of Sociology* 17:62-68.

____. 1993. "The Psychopathology of an Everyday Sociologist: Why Bryant Should Get His Philosophy of Science by Prescription Rather than Off the Shelf." *Canadian Journal of Sociology* 18:65-69.

Gellner, Ernest. (1960) 1973. "Winch's Idea of a Social Science." Pp. 47-49 in *Cause and Meaning in the Social Sciences*. Routledge & Kegan Paul.

____. 1992. *Reason and Culture: The Historic Role of Rationality and Rationalism*. Blackwell.

Giddens, Anthony. 1974. Introduction. Pp. 1-22 in *Positivism and Sociology*, edited by Anthony Giddens. Heinemann.

____. 1984. *The Constitution of Society: Outline of the Theory of Structuration*. University of California Press.

Gitlin, Todd. 1989. "Postmodernism: Roots and Politics." Pp. 347-60 in *Cultural Politics in Contemporary America*, edited by Ian H. Angus and Sut Jhally. Routledge.

Gouldner, Alvin. 1966. *Enter Plato: Classical Greece and the Origin of Social Theory*. Basic.

Gurnah, Ahmed, and Alan Scott. 1992. *The Uncertain Science: Criticism of Sociological Formalism*. Routledge.

Hammersley, Martyn. 1992. *What's Wrong with Ethnography*. Routledge.

Holmwood, John, and Alexander Stewart. 1991. *Explanation and Social Theory*. St. Martin's Press.

Innis, Harold. 1964. *The Bias of Communication*. University of Toronto Press.

Kaplan, Abraham. 1964. *The Conduct of Inquiry*. Chandler.

Kellner, Douglas. 1988. "Postmodernism as Social Theory." *Theory, Culture and Society* 5:239-69.

Kilminster, Richard. 1989. "Sociology and the Professional Culture of Philosophers." Pp. 289-312 in *Social Structure and Culture*, edited by Hans Haferkamp. Walter de Gruyter.

Latour, Bruno, and Steve Woolgar. 1979. *Laboratory Life: The Social Construction of Scientific Facts*. Sage.

Layder, Derek. 1990. *The Realist Image in Sociology*. St. Martin's Press.

Lemert, Charles. 1992. "General Social Theory, Irony, Postmodernism." Pp. 17-46 in *Postmodernism and Social Theory: The Debate over General Theory*, edited by Steven Seidman and David G. Wagner. Blackwell.

Lepenies, Wolf. 1988. *Between Literature and Science: The Rise of Sociology*. Eds. de la Maison des Sciences de l'Homme and Cambridge University Press.

Lichtenberger, James P. 1936. *Development of Social Theory*. Appleton Century.

Lynch, Michael. 1985. *Art and Artifact in Laboratory Science: A Study of Shop Work and Shop Talk in a Research Laboratory*. Routledge & Kegan Paul.

Lyotard, Jean Francois. 1984. *The Postmodern Condition*. (*Theory and History of Literature*. Vol. 10). University of Minnesota Press.

____. [1986] 1989. "Defining the Postmodern." Pp. 7-10 in *Postmodernism: ICA Documents*, edited by Lisa Appignanesi. Free Association Books.

Macklin, Ruth. 1985. "Standards in Philosophy: Pure and Impure? Or Applied and Unapplied?" Pp. 271-87 in *Applying the Humanities*, edited by Daniel Callahan, Arthur Caplan, and Bruce Jennings. Plenum.

McLuhan, Marshall. 1964. *Understanding Media: The Extensions of Man*. New American Library.

Merleau-Ponty, Maurice. 1974. *Phenomenology, Language and Sociology: Selected Essays of Maurice Merleau-Ponty*, edited by John O'Neill. Heinemann.

Merelman, Richard M. 1991. *Partial Visions: Culture and Politics in Britain, Canada, and the United States*. University of Wisconsin Press.

Meyrowitz, Joshua. 1985. *No Sense of Place: The Impact of Electronic Media on Social Behavior*. Oxford University Press.

Misgeld, Dieter. 1986. "Practical Reasoning and Social Science: From the Phenomenology of the Social World to a Radical Hermeneutics." Pp. 71-103 in *European Philosophy and the Human Sciences*, edited by Simon Glynn. Gower.

Montague, William Pepperell. 1927. "The Social Sciences and Philosophy." Pp. 467-83 in *The Social Sciences and Their Interrelations*, edited by William F. Ogburn and Alexander Goldenweiser. Houghton Mifflin.

Morris, Meaghan. 1992. "The Man in the Mirror: David Harvey's 'Condition' of Postmodernity." *Theory, Culture and Society* 9:253-79.

Nicholson, Linda. 1992. "On the Postmodern Barricades: Feminism, Politics, and Theory." Pp. 82-100 in *Postmodernism and Social Theory: The Debate over General Theory*, edited by Steven Seidman and David G. Wagner. Blackwell.

Peters, John Durham, and Eric W. Rothenbuhler. 1989. "The Reality of Construction." Pp. 11-27 in *Rhetoric in the Human Sciences*, edited by Herbert W. Simons. Sage.

Ragin, Charles, and David Zaret. 1983. "Theory and Method in Comparative Research: Two Strategies." *Social Forces* 61:731-54.

Ramsperger, Albert G. 1942. *Philosophies of Science*. Crofts.

Richardson, Laurel. 1993. Review of *Feminist Methods in Social Research* by Shulamit Reinharz with Lynn Davidman. *American Journal of Sociology* 98:913-14.

Rickman, H.P. 1983. *The Adventure of Reason: The Uses of Philosophy in Sociology*. Greenwood.

Rosenau, Pauline. 1992. *Post-Modernism and the Social Sciences: Insights, Inroads, and Intrusions*. Princeton University Press.

Ross, Dorothy. 1991. *The Origins of American Social Science*. Cambridge University Press.

Schutz, Alfred. 1963. "Common-Sense and Scientific Interpretation of Human Action." Pp. 302-46 in *Philosophy of the Social Sciences: A Reader*, edited by Maurice Natanson. Random.

Seidman, Steven, and David G. Wagner (eds.). 1992. *Postmodernism and Social Theory: The Debate over General Theory*. Blackwell.

Selznick, Philip. 1992. *The Moral Commonwealth: Social Theory and the Promise of Community*. University of California Press.

Shalin, Dimitri N. 1992. "Critical Theory and the Pragmatist Challenge." *American Journal of Sociology* 98:237-79.

Skocpol, Theda. 1987. "The Dead End of Metatheory." *Contemporary Sociology* 16:10-12.

Smart, Barry. 1992. *Modern Conditions, Postmodern Controversies*. Routledge.

____. 1993. *Postmodernity*. Routledge.

Smith, Joel. 1991a. "A Methodology for Twenty-First Century Sociology." *Social Forces* 70:1-17.

____. 1991b. "Culture Nationale et Entertainment en Canada: Un Faux Problème?" Pp. 241-63 in *Souveraineté et Protectionisme en Matière Culturelle*, edited by Dave Atkinson, Ivan Bernier, and Florian Savageau. Presses de l'Université du Québec.

Smith, Joel, Robert C. Bealer, and Frank M. Sim. 1962. "Communication and the 'Consequences' of Communication." *Sociological Inquiry* 32:10-15.

Stinchcombe, Arthur. 1993. Review of *John H. Goldthorpe: Consensus and Controversy*. *Social Forces* 71:828-29.

Taylor, Charles. 1989. *Sources of the Self: The Making of Modern Identity*. Harvard University Press.

____. 1990. "Modes of Civil Society." *Public Culture* 3:95-117.

____. 1991. *The Malaise of Modernity*. Anansi.

Thomas, William I. 1951. *Social Behavior and Personality: Contributions of W. I. Thomas to Theory and Social Research*, edited by Edmund H. Volkart. Social Science Research Council.

Turner, Jonathan. 1992. "The Promise of Positivism." Pp. 156-78 in *Postmodernism and Social Theory: The Debate over General Theory*, edited by Steven Seidman and David G. Wagner. Blackwell.

Turner, Stephen. 1992. "The Strange Life and Hard Times of the Concept of General Theory in Sociology: A Short History of Hope." Pp. 101-33 in *Postmodernism and Social Theory: The Debate over General Theory*, edited by Steven Seidman and David G. Wagner. Blackwell.

____. 1994. *The Social Theory of Practices: Tradition, Tacit Knowledge, and Presuppositions*. University of Chicago Press.

Wacquant, Loïc J.D. 1992. Toward a Social Praxeology: The Structure and Logic of Bourdieu's Sociology. Pp. 1-59 in *An Invitation to Reflexive Sociology*, edited by Pierre Bourdieu and Loïc J.D. Wacquant. University of Chicago Press.

Wagner, David G. 1992. "Daring Modesty: On Metatheory, Observation, and Theory Growth." Pp. 199-220 in *Postmodernism and Social Theory: The Debate over General Theory*, edited by Steven Seidman and David G. Wagner. Blackwell.

Wagner, Peter. 1990. "Science of Society Lost: On the Failure to Establish Sociology in Europe during the 'Classical' Period." Pp. 219-45 in *Discourses on Society: The Shaping of the Social Science Disciplines*. Vol. 15, edited by Peter Wagner, Björn Wittrock, and Richard Whitley. Kluwer.

Wolfe, Alan. 1992. "Weak Sociology/Strong Sociologists: Consequences and Contradictions of a Field in Turmoil." *Social Research* 59:759-79.

Wortman, Marc. 1987. Review of *The Post-Modern Aura: The Act of Fiction in an Age of Inflation* by Charles Newman. *Telos* 72:171-78.

Zaret, David. 1980. "Ideology and Organization in Puritanism." *Archives Européenes de Sociologie* 21:83-115.

____. 1985. *The Heavenly Contract: Ideology and Organization in Pre-Revolutionary Puritanism.* University of Chicago Press.

____. 1992. "Religion, Science, and Printing in the Public Spheres in Seventeenth-Century England." Pp. 212-23 in *Habermas and the Public Sphere,* edited by Craig Calhoun. MIT Press.

Zeitlin, Irving. 1968. *Ideology and the Development of Sociological Theory*. Prentice-Hall.

Perverse Effects of Social Support: Publics and Performance in Major League Sports*

ERIC M. LEIFER, *Curriculum Products, Inc.*

Abstract

Random samples have been used to link social support to a diversity of positive outcomes. When viewed within an interaction system, however, social support appears within a context of contention and disserves some while serving others. The manner in which social support is generated can both induce and suppress outcome inequalities. This is shown in the effects of fan support on team performance in major league sports. For the sports most dependent on partisan local publics, fan support functions to suppress the performance inequalities that leagues build into competition. For the sports that have cultivated expectation-driven national publics, fan support functions to induce performance inequalities that leagues seek to eliminate from competition. A computer simulation reveals that the entire amount of actual performance inequality in national public-oriented NFL football can be explained by this perverse effect of supportive publics. Perverse effects, it is argued, are essential to the functioning of interaction systems.

Social support has been linked to a diverse array of positive outcomes, including stress reduction (Haines & Hurlbert 1992; House, Umberson & Landis 1988), decreased mortality and suicide (Cornell 1992; Pescosolido 1990), career mobility (Ibarra 1992), church membership (Bainbridge 1990), and life satisfaction (Stevens 1992). Random samples of individuals have provided the empirical bases for these links, using measurement instruments that transform social support into an individual attribute. When social support is examined within the contexts where it operates, however, its role as a panacea ceases to be so clear. Social support is mobilized when there is contention between individuals and groups and is proffered only at the risk of taking sides. The support one side receives can easily be construed as a lack of support by another. Far from

**Support from the National Science Foundation (SES 8610363) is gratefully acknowledged. The paper has benefited from the comments of Karen Barkey, Ron Breiger, Ronan van Rossem, Wolf Heydebrand, two reviewers, and participants in a Center for Social Science Workshop at Columbia University.* *Direct correspondence to Eric M. Leifer, 114 West 80th St., Suite 1F, New York, NY 10024.*

 Social Forces, September 1995, 74(1):81-118

simply making everything better, social support can function to induce, reinforce, or sometimes diminish various inequalities, and which effects it has depends on where and how it gets cast (Chase 1984; Dekeseredy 1990). Its distributional, as opposed to therapeutic, consequences need to be explored.

Social support gets cast on multiple bases, and as a result has multiple consequences. It may flow from deeply ingrained particularisms, or partisanships, like loyalty to family or tribe or city (White & Riedman 1992). Or it may derive from more universalistic sources, like talent or past performance (Hougland, Billings & Wood 1990). The mix of these bases can be complex. In Whyte's (1946) classic study of the Norton street gang, loyalty to the gang ensured each member some security and stature in the neighborhood. Yet intense focus both within and across rival gangs on performance in a stream of activities, like fighting, baseball, dating, and bowling, served to link support also with performance. In major league sports, the setting used in this article, the partisan support that local publics give a home team has been linked to a widely recognized home advantage factor, which can increase the home team's chances of winning by as much as 30% (Mizruchi 1985; Schwartz & Barsky 1977). Little attention, however, has been given to the different kind of support that came with the mobilization of national publics through televised sports. Instead of loyally following teams as fans, modern sports enthusiasts are drawn into leaguewide competition in the search for winners to support. Partisanship is replaced by expectations regarding performance in determining how support is allocated. This article shows that these different bases for support can have different consequences for the way support affects performance outcomes.

Using game-outcome data, an attempt is made to disentangle the distinct effects of partisan local publics and expectation-driven national publics on performance across the four major league sports. A computer simulation is then used to explore the cumulative consequences of partisan and expectation effects on season performance inequality, to see how much actual performance inequality can be accounted for solely by the social support of publics. The exercise reveals two perverse effects of social support. Where major leagues pursue competitive balance, national publics emerge that can actually sustain performance inequality (Leifer 1990a). Where major leagues seek to reinforce the dominance of large-city teams, the partisan support of local publics promotes performance leveling. In the conclusion, we conjecture that perverse effects of social support may be a general feature of sustainable social systems.

Background and Hypotheses

There is long-standing evidence that supportive publics can favorably affect a diverse array of performances (Allport 1920; Cottrell 1972; Paulus & Cornelius 1974; Travis 1925; Zajonc 1966). In major league sports, teams play half their games in the same "home" location and have long depended on the partisan support of home crowds for both their financial and their competitive performance. In accounting for game outcomes, the *Chicago Tribune* made more references to home and away location than to such factors as talent, injuries, and momentum (Edwards 1979). The actual magnitude of the home advantage

(home win proportion - away win proportion) has been calculated for a diversity of sports settings. Using a composite of prior studies, Courneya and Carron (1992) calculated the home win percentage (HWP) and home advantage (HA) for the four sports (using college and professional versions), which correspond closely to the major league data used in this study, as seen in Table 1. While the home advantage is an accepted fact, its causes and consequences remain somewhat mysterious. We will argue that it derives from a "partisan effect" of supportive local publics.

As Schwartz and Barsky (1977) originally observed, the home advantage is strongest in the indoor sports of hockey and basketball, where partisan home team support can be most intense. In the two predominantly outdoor sports, the stronger home advantage of football seems consistent with its larger and more enveloping crowds. When professional football teams (1960-90) have domed stadiums, the home advantage is .147 compared to .121 for teams with open stadiums. For Major League Baseball, Zeller and Jurkovac (1988) found a home advantage of .105 for teams with domed stadiums compared to .072 for teams with open stadiums, though I found only .089 versus .077 using a longer time frame (from the first domed stadium in 1965 to 1990). To further pursue the apparent importance of crowd intensity, actual crowd size and behavior must be examined.[1]

Within-sport efforts to confirm the role of crowd density have, however, failed to produce clear results across studies (Dowie 1982; Pollard 1986). Density, alone, is not sufficient to boost performance. For example, Thirer and Rampey (1979) studied the effects of spectator behavior on performance in college basketball games and found that antisocial behavior, such as shouting obscenities and throwing objects on the court, negatively affected the home team more than the visiting team by inciting "dysfunctional aggression" by home players. Greer (1983) focused on sustained, but not antisocial, protest among spectators, analyzing the effect of such instances on a number of performance measures. He found that episodes of spectator protest were related to an increase in the performance advantages enjoyed by home teams, using observations from two enthusiastically supported college teams. These findings support a "social reinforcement" theory, where audience behavior interacts with task performance, rather than audiences having effects by their mere presence. Since much of the spectator protest was directed at referees, at least some of the home advantage may have been due to referee intimidation by the home crowd.

McGuire et al. (1992) introduce the reverse causal direction in their study of aggression in 840 hockey games. Finding that high levels of player aggression (measured in terms of penalties) boost the winning chances of the home team, they suggest that aggression is used to entice the crowd into more active involvement from which the home team benefits. Visiting teams struggle to leave the home crowd out of the game by checking the escalation of aggression. Since aggressive penalties are remarkably balanced between home and visiting teams, game outcomes are significantly affected by the strategic struggle over mutual aggression levels. This intriguing finding might help account for the across-sport usefulness of crowd intensity in explaining overall home advantage differences, and within-sport difficulties in linking crowd density to any home

TABLE 1: Home Advantage Levels across Sports, in a Composite of Past Studies and in the Present Study

	Composite Home Advantage Results[a]			
Sport	Studies	Games	Home Win Percent	Home Advantage
Baseball	6	23,034	53.5	.070
Football	5	2,592	57.3	.146
Hockey	4	4,322	61.1	.222
Basketball	8	13,596	64.4	.288
		Results for Extant U.S. Major Leagues		
Sport (Seasons)		Team-Seasons	Standard Deviation	Home Advantage
Baseball (1901-90)		1,672	.084	.085
Football (1960-90)		808	.232	.124
Hockey (1917-89)		671	.117	.196
Basketball (1949-89)		636	.129	.290

[a] Source: Courneya & Carron (1992)

advantage. Whether crowds of varying size are activated to produce an effect depends on the dynamics of the competition on the field.

Performance effects of publics cannot be understood in terms of attributes of either publics or competitors when viewed separately. Some researchers, for example, have tried to resolve whether the home team performs better or the away team performs worse before a partisan home crowd. Schwartz and Barsky (1977) looked at both defensive and offensive performance measures in Major League Baseball (such as errors, double plays, runs per at bat) and claimed the home advantage inheres exclusively in offensive superiority after finding the home team had higher levels on offensive performance measures. Andrew (1984) looked at five performance measures in Atlantic Coast Conference Basketball and found significant differences between home and away performance measures in each measure except free-throw percentage. He then used the coaches' "expected" performance levels (in a neutral location?) to determine whether the observed measures were high at home or just low away, claiming support for the latter.

These efforts ignore crucial interdependencies in the sports setting. One team's win is another's loss. A great defensive performance is facilitated by a poor offensive one, and vice versa. Home advantage research often tries to get two pieces of information out of a single performance measure — how "good" each side was in producing it. Instead, publics should be thought of as affecting the nature of the competition between teams rather than the individual playing

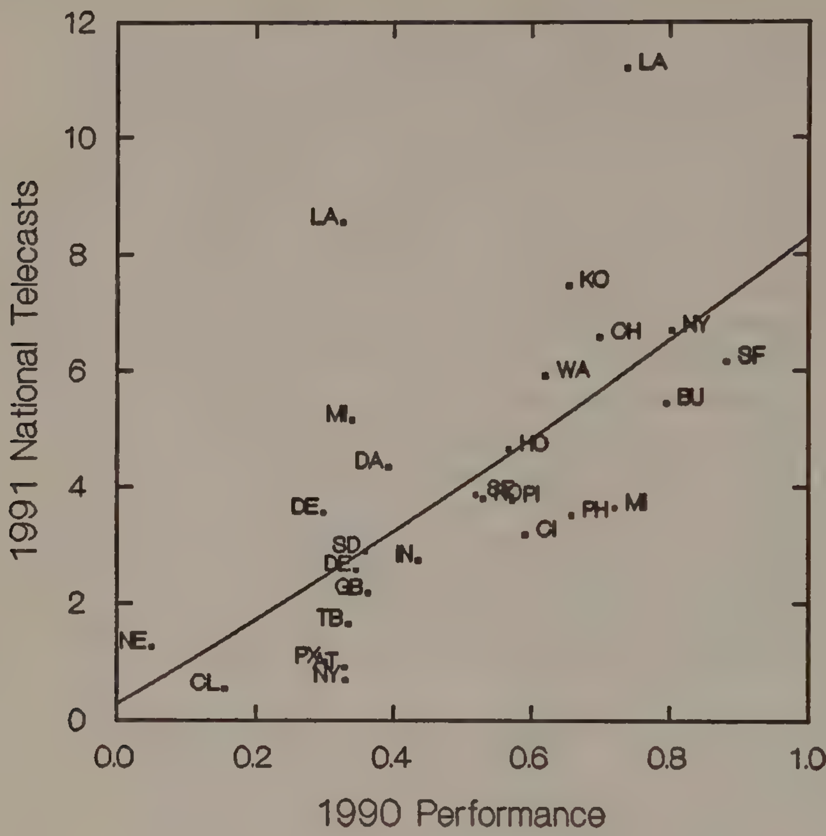
qualities of the teams. Publics don't make teams better or worse; instead, they define the framework in which competition is interpreted and in so doing affect the course of competition itself. This allows the game to be the most basic unit, rather than teams, and allows a clear separation of the home advantage from playing qualities of teams. With this change of imagery, it becomes meaningless to ask which team is affected by a public. Competition is affected.

What interpretive framework is imposed on competition depends on how publics are organized. Home advantage research has acknowledged only one type of public, the partisan local public. Though they may vary in size, involvement, or courtesy, the partisan local public provides unconditional support for the home team. The fact that in every major league sport half of each team's games are played in the same "home" location, before crowds that certainly appear supportive, underscores the importance of partisan local publics. Yet most of those who show up at games, and hordes that do not, are increasingly part of national publics insofar as they tune in to (on television and in print) the same games and analyses as the rest of the nation. Not bound by place, they follow leaguewide competition in search of winners to support. At a game, they can be as familiar with the away team as with the home team and heavily swayed by expectations in their orientation to the outcome. Persisting localism may lead them to want the home team to win, but their grasp of leaguewide competition may lead them to expect the away team to win. This, I argue, will affect the way they behave as publics and their impact on outcomes.

Both local and national publics are creations of leagues, first in attaching teams to cities and then in attaching major league sports to network television. The impact of publics on performance must be viewed in relation to the organizing aims of leagues. The pursuit of local publics created an incentive within leagues to ensure that large-city teams dominated competition, as winning was worth most in the largest local markets (Quirk & Fort 1992). With teams strongly attached to cities, the viability of smaller-city teams has often hinged on the loyalty of partisan local publics persisting despite poor winning chances. Once network television made national publics possible, however, the location of winning teams diminished in importance. Instead of cultivating fan loyalty to a single (home) team, leagues had to lure fans into following leaguewide competition so that fans everywhere might tune into a few selected telecast games. The possibility of national publics made competitive equality a crucial organizational aim for leagues, as it lured fans to search across an entire league for winners to support (Leifer 1995). At the same time, it made winning an imperative for any team seeking to mobilize a national public (Klatell & Marcus 1988; Rader 1984).

As leagues organize competition to accommodate distinct publics, the latitude for these publics to affect competition gets channeled in distinct directions. The loyalty and partisanship of local publics is channeled toward supporting "their" teams. One often hears a sense of guilt among fans if they fail to watch or attend a game involving "their" team, as if their absence could somehow be responsible for the team's loss. Support becomes a show of strength, like foot soldiers behind a mounted commander. If it actually helps the home team win, it would function as a performance leveler, for half of each

FIGURE 1: Relation between the Number of Nationally Televised NFL Games and the Prior Season Performance of Scheduled Teams[a]

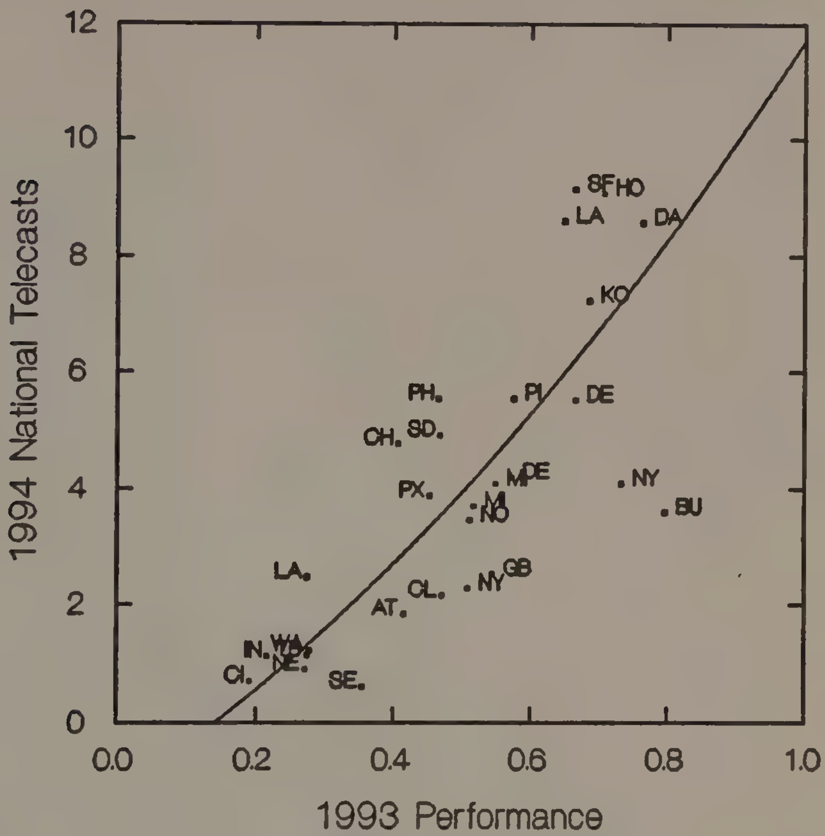


[a] Random noise was introduced to scatter overlapping observations.

team's games are played at home (and half away, where home support for the opponent works against them). Strong teams would be reined in during away games, and weak teams would be boosted during home games. In this way, local publics work against league designs for large-city dominance by hindering strong and weak teams from being strong and weak, respectively, in half the games they play (perverse effect 1).

National publics are not constrained by fixed loyalties. Instead, they cast their support on the basis of performance expectations formed from following leaguewide competition. Past success becomes the most reliable basis for molding these performance expectations. Winning teams amass followings, and losing teams are inevitably abandoned. Winning teams are less supported than adulated, and fans are less foot soldiers than onlookers eager to be associated with the winning side. As can be seen in Figure 1, NFL teams that win in a given season get more national television exposure in the next season. They also get more positive exposure in print and more support from enthusiasts they encounter. Insofar as high (low) performance expectations help a team continue

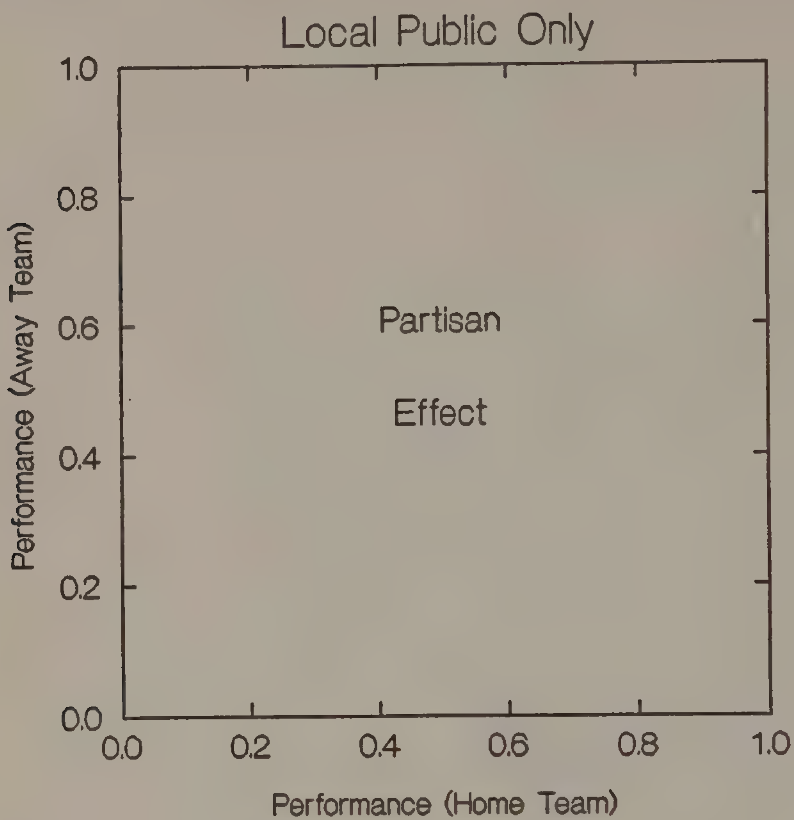
FIGURE 1: Relation between the Number of Nationally Televised NFL Games and the Prior Season Performance of Scheduled Teams[a] (Continued)



winning (losing), they function to induce performance inequalities in what is widely known as a "Matthew effect" (Merton 1957; Podolny 1992). The more successful leagues are in pursuing competitive equality, the more latitude a national public has in inducing performance inequalities, and the more energy it puts into affecting performance outcomes in this perverse manner (perverse effect 2). Real inequality would shut off the national public's influence on and interest in performance outcomes.

In this scenario, publics do not merely follow what happens. They mobilize and react on the basis of an interpretive framework. Where local publics predominate, the failings of a team cannot be divorced from the failings of its local supporters. The show of partisan support among local publics can become as important as the winning and losing of teams. Only where national publics predominate is a relentless zero-sum framework imposed on competition. Every play, every game, and every series is used to sort and resort teams into winner and loser roles. A bit of chance, like a favorable gust of wind or a lucky bounce,

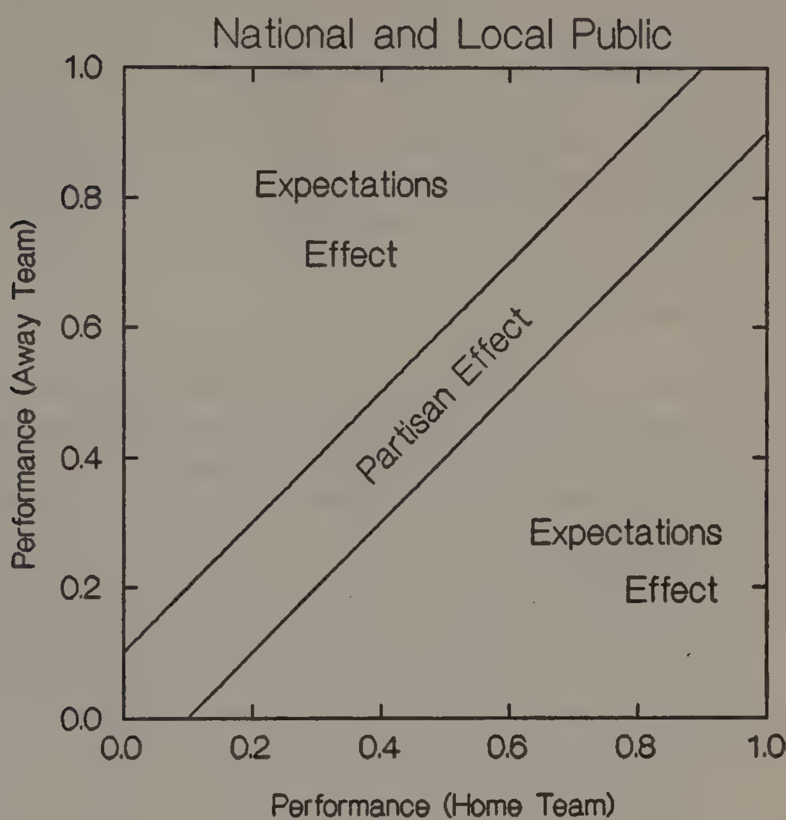
FIGURE 2: Predicted Effects of Publics When Different Kinds of Publics Are Taken into Account



is enough to make two similar teams very different. In the winning locker room after a championship game, it is completely forgotten that only minutes earlier either team could have won.

Major league competition is not intrinsically zero-sum. It can even seem odd that winning and losing are so important in a setting where teams are so much alike. Imagine a world where the public appreciated the competitive balance of teams, getting enjoyment out of the delicate balance maintained by evenly matched teams. The events that led to the breakdown of this balance, that produced a winner and loser, would be viewed as flaws that disrupted the delicate coordination and would evoke disappointment rather than cheers. In this world, public pressure would be directed toward suppressing the emergence of winners and losers, as both would be equated with failure. If we (or major league organizers) look at what fans actually do, instead of their interpretations, this alternative world is not so strange. Fans are quick to leave sporting contests early, or cease following pennant races, when the delicate balance is broken, whether the team they supported is winning or losing. The modern insistence on finding winners and losers only remains viable if leagues work hard to maintain competitive equality. Far from being natural, the zero-sum context national publics impose on competition is antithetical to the basis on which competition is organized.

FIGURE 2: Predicted Effects of Publics When Different Kinds of Publics Are Taken into Account (Continued)



Interpretive context affects competition and not the intrinsic abilities of either home or away teams. Effects will depend on pairings, not on teams. Think of these effects in terms of the performance box with home and away team performance on the axes, as shown in Figure 2. If all publics were local and support for the home team were cast regardless of past performance, then the partisan effect should be uniform throughout the performance box. This is the implicit prediction of most home advantage studies. Where national publics have emerged and influence the orientation of home crowds, the partisan effect should be largely confined to the diagonal of the performance box where no clear expectations regarding the outcome are possible. Off the diagonal, even otherwise partisan home crowds will subtly implicate themselves in outcomes that are perceived as inevitable.[2] As a result, the partisan effect will be diminished and teams will tend to perform as they are expected. In the next section, a performance box is used to examine nonuniformities in the observed home advantage across performance pairings.

The major league sports of football, baseball, basketball, and hockey vary considerably in the extent to which they have been able to mobilize expectation-driven national publics in the order listed (Leifer 1995). Network television broadcasts (for regular season games) are the best indicator for success here. Nearly 70% of National Football League (NFL) team revenues come from an

equally shared network television contract. With a doubleheader on Sunday and a Monday night game, football fans everywhere can tune in to the same carefully selected games. At the other extreme, National Hockey League (NHL) teams depend almost entirely on local box-office receipts and local (and cable) broadcasting rights to survive, ensuring a highly partisan setting (and limited public) for each game. Major League Baseball (MLB) and the National Basketball Association (NBA) fall in between these extremes, with MLB deriving more revenue from network broadcasts than the NBA in the year studied (1980). The differences between sports in the way support is cast are great enough to find distinct consequences of support on performance.

Data and Estimates

Game-level attendance and outcome data was gathered for the 1980 major league season in baseball (2,092 games), basketball (942 games), and hockey (836 games), and, because of the short 16-game season, the 1974-80 seasons for football (1,428 games). The 1980 focal season was selected because it was not disrupted by labor disputes in any of the sports. It also was during the pinnacle of network domination of sport broadcasting, before the explosion of cable, satellite, and pay-per-view options that have since muddied the distinction between local and national publics. Overall, the home advantage was .078 for baseball, .136 for football, .192 for hockey, and .236 for basketball, making the 1980 focal season reflective of the long-term differences between the sports observed earlier.

Game outcomes were used to compute a running measure of performance that corresponds to season standings given in newspapers, from which fans follow the progress of teams throughout the sport season. Game attendance figures (reported in the *Los Angeles Times*) were used to compute crowd densities, or the proportion of each stadium or arena that was filled for each game.[3] To assess the role of performance for mobilizing crowds, particularities of each sport loom large. Baseball teams play 81 home games a season, during days and evenings of every day of the week. Even the most enthusiastic fans must be selective in the games they attend. At the opposite extreme, football teams play only eight home games a season, almost all conveniently scheduled on Sundays. Football often can fill stadiums with season-ticket holders, who commit themselves before the season even begins. To complicate matters, baseball and football teams sometimes share the same stadium, as do basketball and hockey teams, so that stadium capacities often bear little relation to actual ticket demand.

In Table 2, average crowd densities and attendance figures are given for each sport. From these figures alone, it can be expected that football and hockey will show less sensitivity to variations in performance than will basketball and particularly baseball, as stadium size sharply limits the number of people that might otherwise attend games. This sharply truncates the variation in turnout that the various performance combinations would otherwise elicit.

Despite these caveats, home crowds remain sensitive to performance in every sport, as is evident in Figure 3. Categories in Figure 3 are defined by the

TABLE 2: Average Attendances and Crowd Densities for the Four Major League Sports

Sport (Season)	Crowd Density	Attendance	Games (Missing)[a]
Football (1974-80)	.822	54,152	1423 (5)
Baseball (1980)	.414	21,596	2086 (6)
Basketball (1980)	.619	10,005	941 (1)
Hockey (1980)	.840	13,445	834 (2)

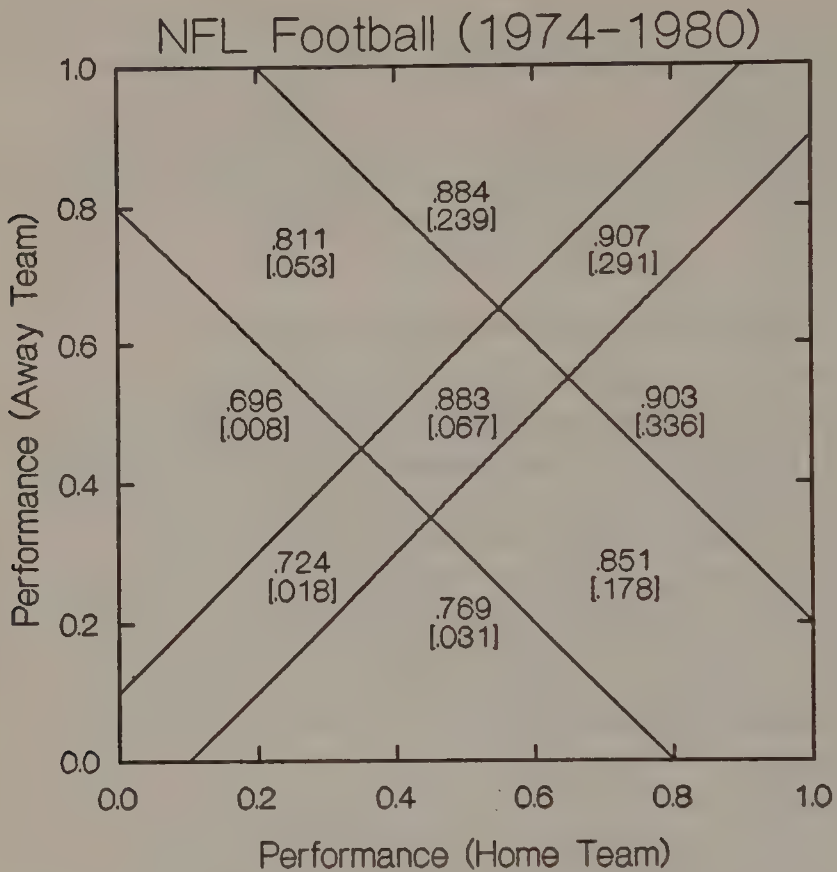
[a] Season games for which no attendance data were available.

relation between the performance levels of home and away teams at the time of the game. From the lower left to upper right of each performance box lie contests where home and away teams had similar performance levels and hence no clear expectations for the outcome were possible. Crowd densities were the highest where both home and away teams were high performers, in every sport. As overall performance levels decline, similarity ceases to attract large publics relative to overall sport averages. In general, local publics are more sensitive to home team performance levels than away team levels, as evidenced from the higher densities in lower right cells than in upper left cells. Proportions of sellouts (density > .98) are given in brackets. These are mostly consistent with crowd density figures.[4]

OLS regression was used to confirm that the pattern in Figure 3 does not depend on the arbitrary categorization of performance or peculiarities in the distribution of stadium sizes. Home and away team performance and their multiplicative interaction, along with stadium size as a control, were introduced into the equation. When regressed on crowd densities, the interaction coefficients were negative in the sports with the highest overall crowd densities. As home and away performance levels rise, increasing crowd density, the capacity cap increasingly comes into play. Where this cap rarely comes into play, as in baseball, the interaction coefficient is strongly positive, indicating that games between high-performing teams have an extraordinary appeal. When the same variables were regressed on the sellout dummy, the interaction coefficients became positive in all sports and significantly so in all but hockey. In no sport except football is the isolated impact of either home or away performance level statistically significant. Nor are intercepts significantly different from zero except in hockey, where sellouts can occur between even the lowest-performing teams. These findings strongly suggest that attendance potential exceeds stadium capacity for games between winning teams outside of baseball.

The same performance boxes can be used to explore the effects of the variable publics on performance. In the context of particular regions of the performance box, the home advantage becomes the difference between home and away winning chances for the same performance combinations at home and away. Home advantage is necessarily symmetrical around the equal perfor-

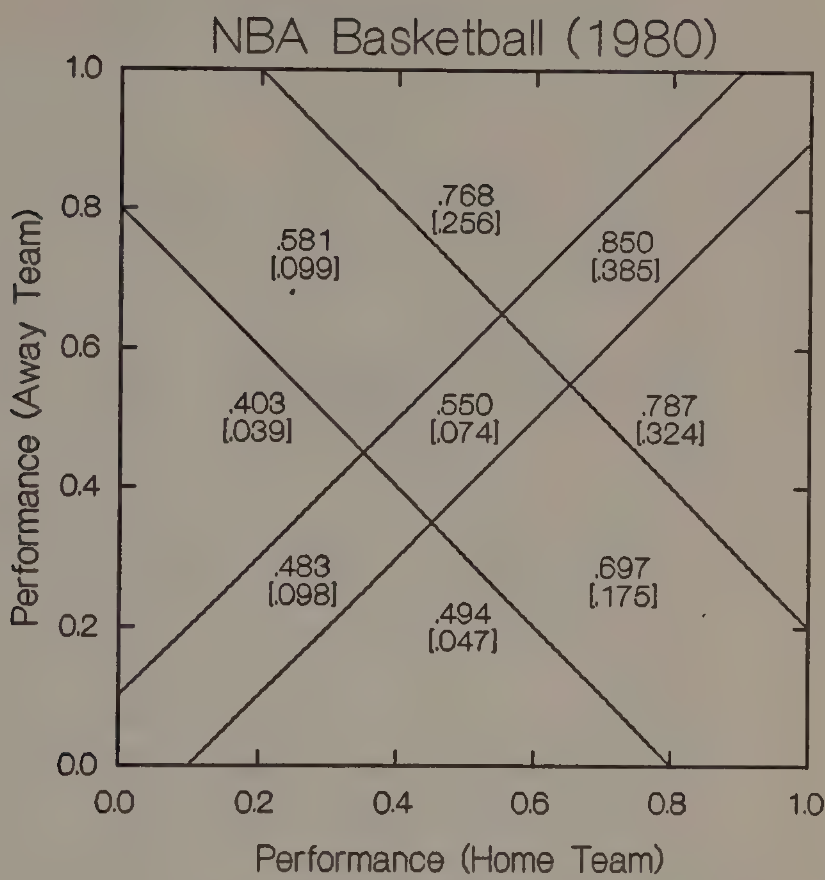
FIGURE 3A: The Effect of Home and Away Team Performance Levels on Average Crowd Density and Proportion of Sellouts[a]



[a] A stricter criterion for similar performance was used for baseball because of this sport's lower level of dispersion in performance.

mance level diagonal, as outcomes on both sides must be used in its calculation. Using a minimum of regions to stabilize estimates of winning chances, Figure 4 shows how dramatically home advantage can vary across performance combinations. Between-sport contrasts are even more striking. Compare, for example, the most nationally televised sport of football with hockey, the sport most lagging in the cultivation of national publics. In football, home advantage is relatively low in games between performance unequals. Among performance equals, home crowds have a sizable impact on helping the home team win and this impact sharply increases as the performance level of both teams rises. Thus there appears to be a parallel between crowd density and home advantage, though the dense crowds that follow winning home teams have little added impact when the home team plays a weaker opponent. As shown in the discussion, stronger teams have similarly high prospects for winning wherever the game is played. Performance expectations prevail.

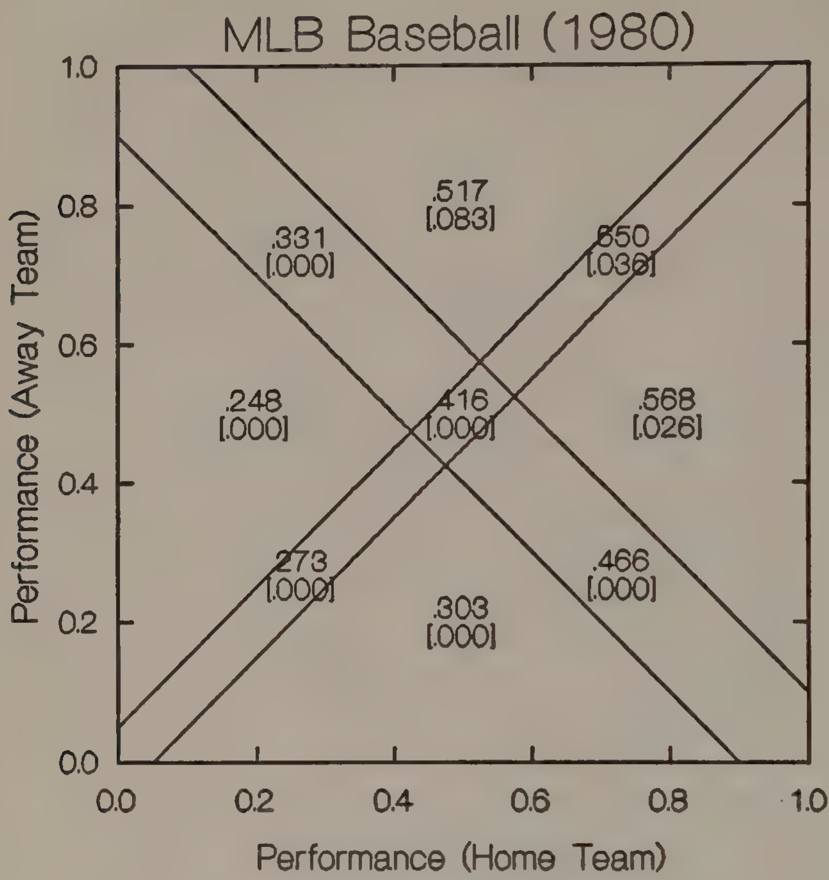
FIGURE 3A: The Effect of Home and Away Team Performance Levels on Average Crowd Density and Proportion of Sellouts[a] (Continued)



Hockey displays virtually the opposite home advantage pattern. Home advantage disappears entirely in games between high performers and increases as performance levels both decrease. In matches between performance unequals, home advantage is relatively high. Though crowds were least sensitive to performance levels in hockey, the fact that a high overall home advantage should disappear where crowds are densest, and appear where they are sparsest (72.3% of capacity) is indeed curious. As developed further below, however, the result is not without function. Local publics, which predominate in hockey, are sensitive to where games are won. Losing teams can maintain local support by winning, albeit rarely, at home. Losses on the road are less debilitating than ones at home, as local publics have themselves to blame for being absent.

Baseball and basketball exhibit curious amalgamations of these extreme patterns. Basketball closely resembles football in games between performance equals and resembles hockey in games between performance unequals. This combination displays the closest parallel with crowd density results. That is, dense crowds appear to magnify home advantage in basketball. For baseball, however, the converse pattern of resemblances hold. Low performers exhibit the

FIGURE 3B: The Effect of Home and Away Team Performance Levels on Average Crowd Density and Proportion of Sellouts[a]

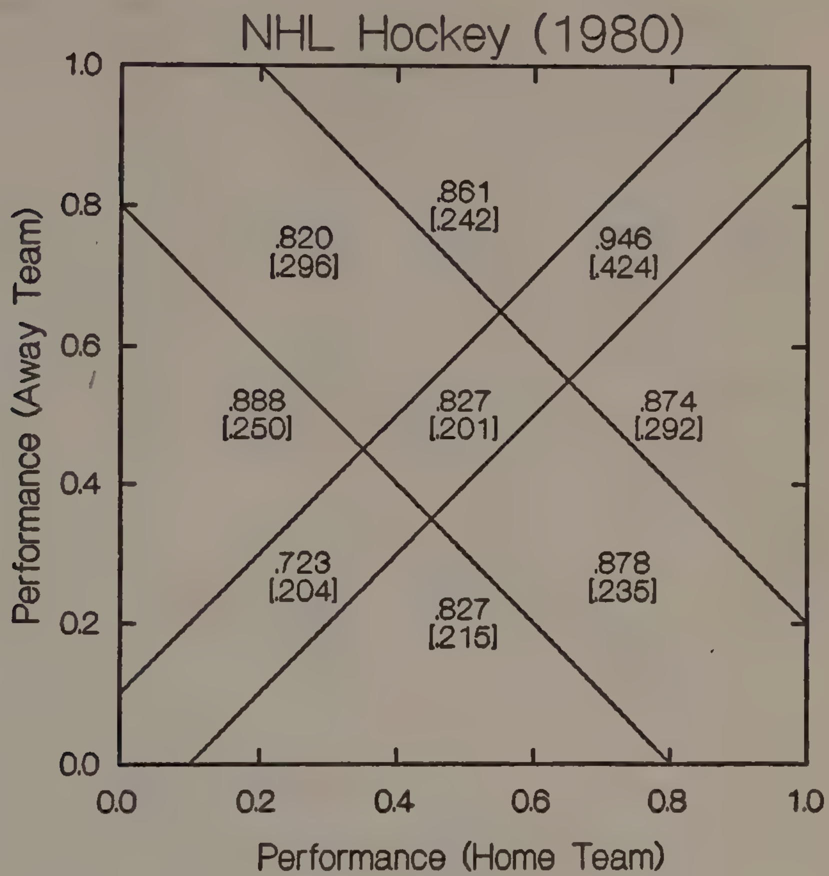


highest home advantage, as in hockey, but unequal performers exhibit relatively low home advantage, as in football. Aside from the slight rise in home advantage in games between high performers, the baseball results are more damning than hockey's for a straightforward crowd density hypothesis, since crowd density was more highly sensitive to performance levels in baseball.[5] Home advantage appears in performance regions where crowds are sparsest in baseball, in sharp contrast with basketball.

Basketball followed a developmental path similar to that of football, remaining primarily a college sport until the modern era, though basketball fell far short of football in cultivating a national public for its regular season play. Baseball shares with hockey a similar developmental path and a continuing dependence on local publics for support, yet it has made more substantial inroads in the creation of national publics. Before trying to link the distinctive role of publics in each sport to its distinctive home advantage pattern, the results need to be further elaborated.

Patterns in Figure 4 are tied to an artificial performance categorization, with some category results based on a relatively small number of observations. As in the crowd density case, regression analysis can be used to test for an underlying

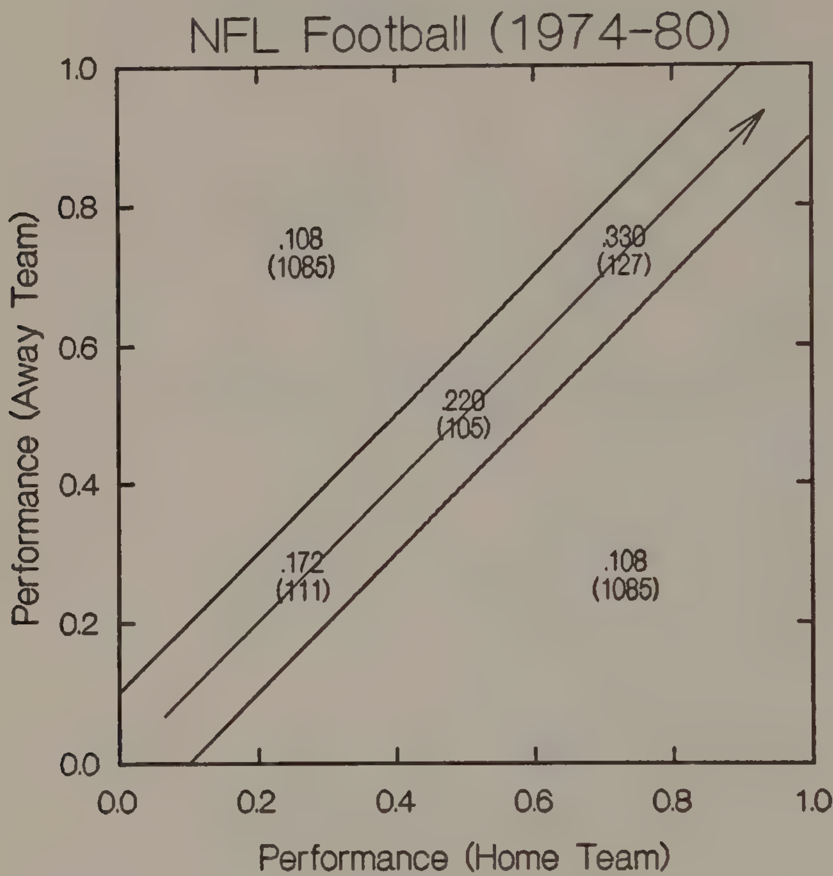
FIGURE 3B: The Effect of Home and Away Team Performance Levels on Average Crowd Density and Proportion of Sellouts[a] (Continued)



logic. In addition, it is necessary to link crowd density, or the lack of it, directly to home advantage patterns. Though in basketball home advantage was the highest in performance regions with the highest crowd densities, for example, the link has not been established at the level of individual games. In any given performance region, actual crowd densities need not be related to home team chances of winning.

To address these limitations, it is convenient to use home team winning chances (HWC) as the dependent variable.[6] HWC is a function of home and away team performance differences, factors tied to the home advantage, and the residual factors of chance and accident. The importance of performance differences can be interpreted in two ways. They may simply be an indicator of skill differences, and hence teams with higher past performance should beat teams with lower past performance. Yet differences would exist even if there were competitive equality and could operate on HWC entirely through the expectations they arouse in both teams and publics. These expectations may subtly modify crowd behavior and the disposition of teams to respond to this behavior. Specifically, national publics should magnify the role of expectations, as they allow visiting teams to be both known and admired by home crowds.

FIGURE 4A: The Effect of Home and Away Team Performance Levels on the Home Advantage. Results Are Symmetrical Around the Similar Performance Axis

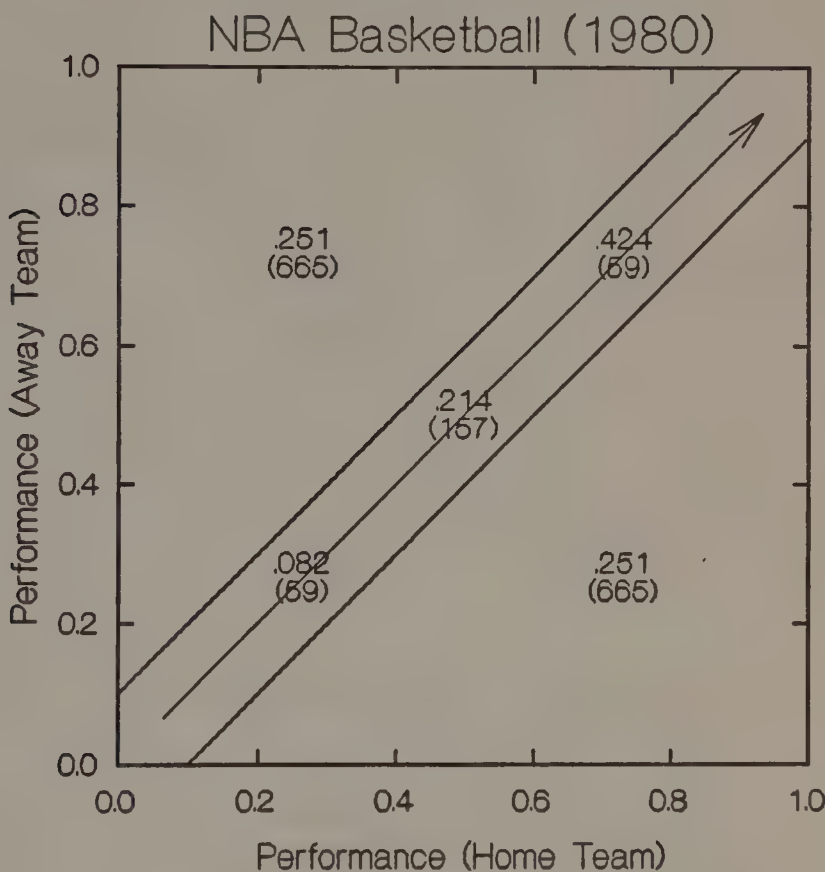


[a] The number of games used to calculate home advantages is given in parentheses.

Hence the fact that higher performers tend to beat lower performers may have as much to do with expectations as it does with skill.

Crowd density and stadium capacity were used as determining factors of home advantage. Since the latter does not vary across games (within the 1980 season) it is only useful for addressing between-team variation. Holding density constant, increases in stadium capacity increase both the absolute size of the crowd and the average distance of the crowd from the competition. These can have opposite impacts on HWC. Yet between-sport differences in home advantage suggest the distance factor may be more important, as sports with the highest home advantage have the smallest and most intimate stadiums. According to the literature reviewed earlier (Schwartz & Barsky 1977), crowd density is supposed to increase HWC when stadium capacity is held constant. To facilitate interpretation, stadium capacities were divided by the largest stadium in each sport, so that the range of both variables falls between zero and

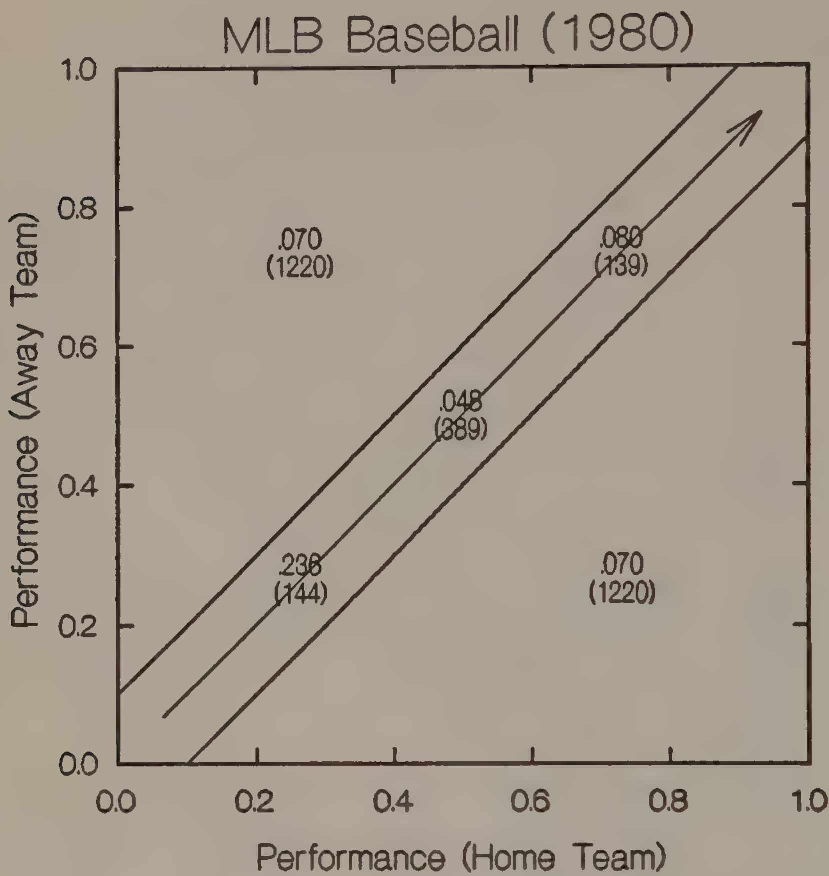
FIGURE 4A: The Effect of Home and Away Team Performance Levels on the Home Advantage. Results Are Symmetrical Around the Similar Performance Axis (Continued)



one.[7] A multiplicative interaction was also constructed, allowing for large, full stadiums to have extra impact, but this was not significant except where noted below.

A two-step analysis was carried out. First, season performance differences (home minus away) at the time of each game were regressed on game outcomes (0 = lose, 0.5 = tie, 1 = win). As shown in Figure 5, the resulting intercepts mirror the overall home win proportions of each sport, which, when differenced from .5, are one-half the overall home advantage. Performance differences show considerable variation in their impact on HWC. In baseball, for example, a home team with an extreme performance advantage of .4 (e.g., .7 to .3) has only a .108 (.4 x .271) better chance of winning than it would playing a team at its own performance level. In basketball, the home team would have a .287 (.4 x .718) better chance in these same circumstances. The overall usefulness of performance differences in predicting game outcomes is measured by the proportion

FIGURE 4B: The Effect of Home and Away Team Performance Levels on the Home Advantage[a]

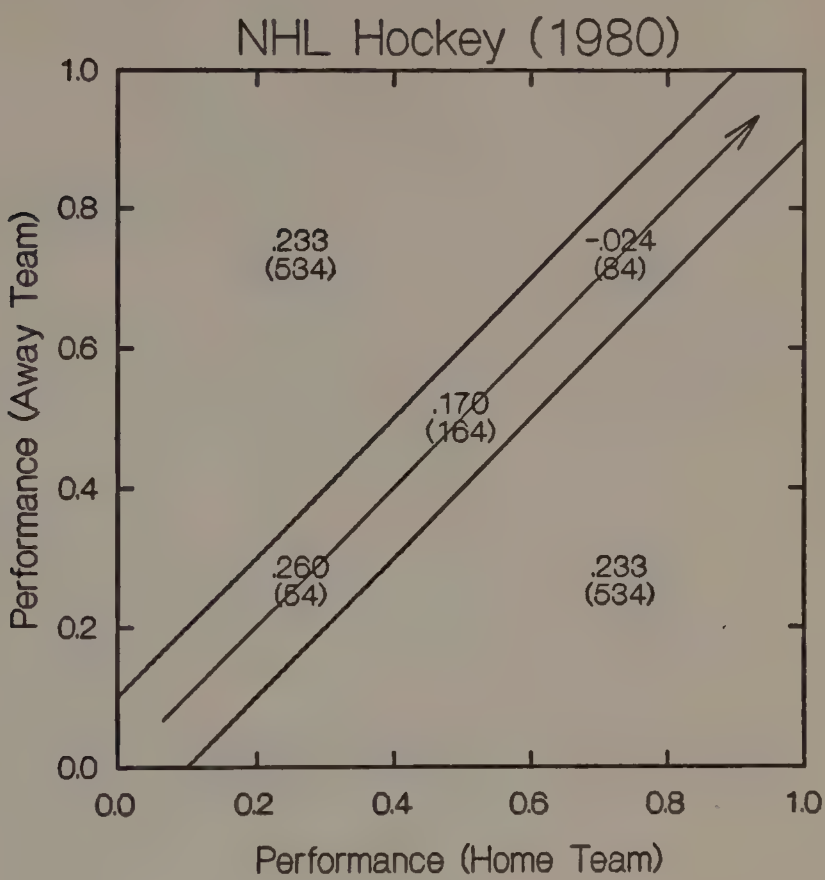


[a] Results are symmetrical around the similar performance axis.

of variation in outcomes accounted for by performance differences, or R^2. This measure varies from .005 in baseball and .078 in hockey to .123 in football and .141 in basketball. If performance were no more than a measure of skill, sports with longer seasons should exhibit higher R^2s, since the measure would on average be based on a larger sample. Yet the results go in the opposite direction, with football teams playing less than one-tenth the number of games that baseball teams do.[8]

Crowd density and stadium capacity were added to the regression equation to account for the residuals.[9] Though the added impact of these variables on HWC was weak, the direction of the impact reinforces the emerging paradox. In football and basketball, home teams had greater chances of winning when denser crowds turned out to support them. Yet the opposite was the case for baseball and hockey, where home teams performed best before the sparsest crowds! In every sport except baseball, larger stadiums decreased home winning chances. A multiplicative interaction term between crowd density and stadium capacity proved insignificant in every sport except baseball, where its inclusion

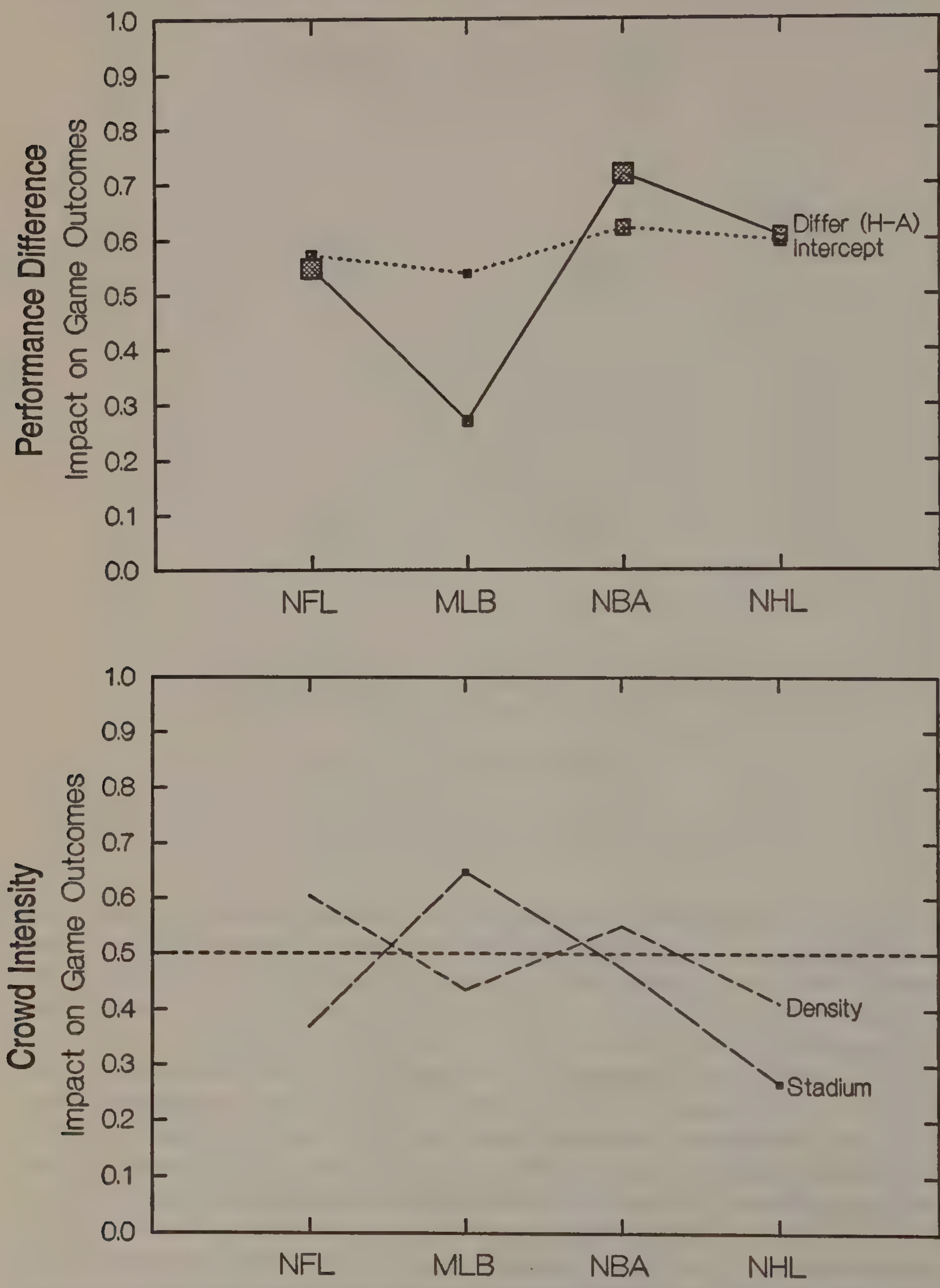
FIGURE 4B: The Effect of Home and Away Team Performance Levels on the Home Advantage[a] (Continued)



both changed the direction of the impact of stadium capacity (to -.157) and made the impact of crowd capacity (-.549) significantly negative. The interaction term, however, had a significant positive impact on HWC (.693), though it was not strong enough to overcome the negative effects of each of its separate components.

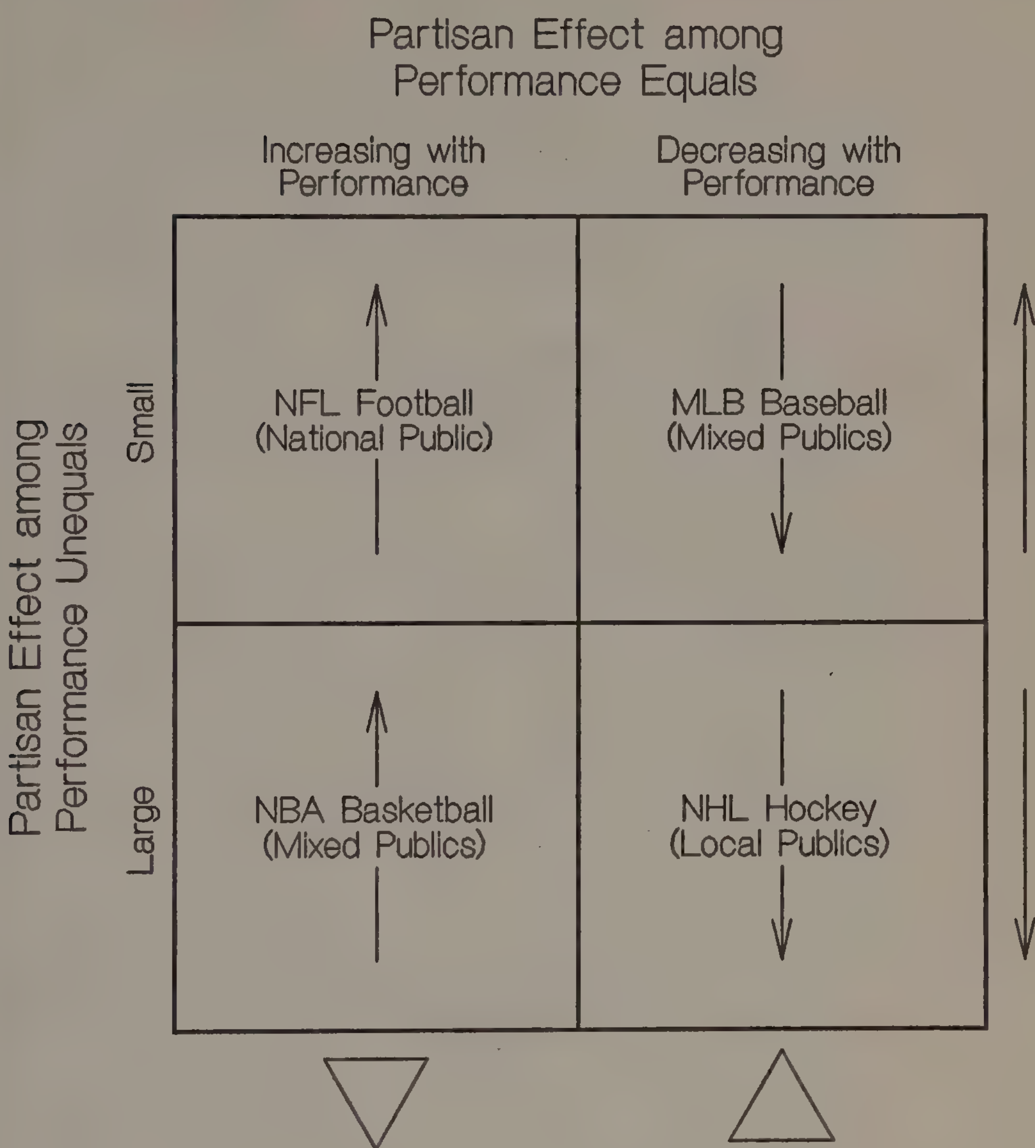
The consistently negative impact of stadium capacity within each sport accords with between-sport differences in overall home advantage. Where stadiums are large and fans quite distant from the competition, as in baseball and football, overall home advantage is lower than in the smaller, more intimate stadiums of hockey and basketball. Stadium variation within each sport is small relative to stadium differences between the sports, and so the fact that within-sport variation still was able to reveal a consistent impact is quite noteworthy. Yet the finding that in baseball and hockey denser crowds of supportive fans can actually lower home winning chances poses a challenge to any straightforward interpretation of the home advantage.

FIGURE 5: Impact Coefficients of Performance Differences and Crowd Intensity Indicators on Game Outcomes[a]



[a] Performance differences were controlled for in the bottom panel results. Size of box denotes degree of statistical significance.
0 = win .5 = tie 1 = win

FIGURE 6: **Summary of Main Findings**[a]



[a] Major leagues differ in the size of the partisan effect for performance equals (columns) and unequals (rows), and in the size of the expectations effect for performance unequals as indicated by arrows through each sport (upward arrows denote a large effect). Since the expectation effect helps magnify performance inequality, and partisan effect lowers it, arrows (in cells and margins) also denote impact on performance inequality. Column symbols indicate whether the column effect promotes performance leveling at the top or bottom of the performance distribution.

Further analysis reveals that crowd density has a different impact that depends on the performance levels of competitors.[10] For both baseball and hockey, crowd density and stadium capacity had their strongest negative, and now significant, impact on HWCs in contests involving performance equals. Among performance equals in football and basketball, large stadiums and dense crowds boosted (significantly in basketball) HWCs (though in football crowds provided the strongest boost for home teams playing higher-performing away teams). In basketball, the impact of crowd density was actually weakly negative in games between performance unequals, confining the positive overall effect entirely to performance equals. Hence crowds tend to have their greatest impact in games between performance equals, though this impact can be channeled in opposite directions across sports.

Figure 6 summarizes the findings in this section. Home advantage patterns reduce neatly to two binary dimensions, with the four sports filling up the four possibilities. Among performance unequals, home advantage was found to be either low or high. Among performance equals, home advantage either increased or decreased with performance levels. Placement on these dimensions appears related to the predominance of national publics in a sport. Football stands in sharpest contrast to hockey with regard to publics, with baseball and basketball treading a middle ground in relying on both local and national publics. Hence the rise of national publics must be associated with low home advantage among performance unequals and increasing home advantage over performance levels among performance equals, as found in football. A third dimension, the impact of performance differences, drives outcomes where home advantage is low. In Figure 6, upward arrows through a sport denote relatively strong impacts and downward arrows denote weak impacts (assessed in terms of R^2s).

All three dimensions have implications for (re)generating patterns of performance inequality. Partisan effects, wherever they appear, function as performance levelers. Hence small partisan effects among performance unequals open the way for the regeneration of performance inequality. Among performance equals, an increasing partisan effect over performance levels will truncate the upper end of the performance distribution, yielding a small group of similarly performing teams at the top (with the opposite pattern yielding a similar group at the bottom of the performance hierarchy). Insofar as performance differences have an impact, they will function to generate performance inequality as winners tend to continue winning and losers to continue losing. The use of up-down arrows (and triangles) in Figure 6 is in reference to these impacts on performance inequality.

In football, for example, winners will tend to split their games among each other, yet for the bulk of games where they play lesser-performing teams they will tend to win wherever the game is played. Hence performance inequality is regenerated among performance unequals yet is clearly capped, as no team will happen to win all its games against other high performers. No such clear cap is induced at the bottom of the performance hierarchy, so some low performers may lose all their games against other low performers. In hockey, these processes are entirely turned around. Low and unequal performers tend to split their games with each other and hence are subject to leveling. Among low

performers in hockey, tendencies to win at home (and reserve losing for away games) may account for the unusually high fan support accorded losing teams. Performance inequality must be generated among higher performers, who win and lose from each other randomly. The resulting lack of close races at the top can serve as an obstacle to cultivating national publics.

Basketball and baseball display peculiar mixtures of effects that represent built-in tensions. Basketball follows football closely, yet both a partisan effect and performance differences effect outcomes among performance unequals. High performers may win nearly all their home games but not much more than half their away games. Performance inequality is generated from the impact of performance differences and suppressed by the partisan effect. For a long time, the bulk of national attention has been focused on two teams, the Boston Celtics and Los Angeles Lakers, and here mainly in the playoffs. If these teams falter, as they have in recent years, the large partisan effect among performance unequals will hinder basketball's ability to produce a new set of nationally hailed winners. In this respect, basketball's situation is more precarious than football's.

Baseball resembles hockey, except it shares with football the weakening of home advantage among performance unequals. Yet performance differences have only a weak impact, so that mostly random inequality is generated from games between performance unequals, as it is in games between equally high performers because of the weakness of the partisan effect. The appearance of a partisan effect among equally low performers functions to truncate the bottom of the performance distribution, yet it is not as effective in bringing out local fans as it is in hockey. Nascent national publics in baseball make overall winning too important for local fans to be satisfied with merely home game wins, yet the predominant localism undermines the impact of performance differences in generating stable winners and losers. Baseball seems to be in the most precarious position of all sports, anchored more by its traditional popularity than by its ability to produce a small group of rival superteams for national publics or a pattern of home game winning that draws local publics.

Simulation Results

Individual games that make up season records are swayed by stellar performances, player injuries, coaching innovations, and a host of other factors we have given no attention to. If in the long run publics have the impact observed in the last section, however, this impact should at least apply pressure for particular patterns of performance inequality to emerge. To isolate these patterns in the clearest possible terms, a computer simulation is used that "plays out" seasons of competition solely on the basis on the effects observed in the last section. These patterns will then be compared to actual patterns, to see how much of observed inequality can be attributed to the complex workings of the partisan and expectation effects.

Driving the simulation is a game outcome equation cast in terms of home winning chances (HWC). To simplify, stadium capacity is assumed to be constant within each sport (and large enough to accommodate ticket demand),

and crowd density is assumed to be entirely a function of contending team performance levels. In addition, a logistic equation is used to constrain HWC between 0 and 1. The resulting equation is:

$$\ln\left(\frac{HWC}{1-HWC}\right) = b_0 + b_1(PFH-PFA) + b_2(equal) + b_3(PFH+PFA)\ x\ equal$$

where

PFH (PFA) = performance level of home (away) team.
equal = a variable equal to 1 if teams have similar (≤.1) performance levels and 0 if not.[11]

In this formulation, b_0 is the partisan effect (i.e., home advantage) for unequal performing contenders, b_1 is the expectations effect, b_2 is the difference in partisan effect between performance equals and unequals (for contending teams with no past wins), and b_3 is the impact on the partisan effect of increasing performance levels among equals. Thus the game outcome equation allows for the effects observed in Figure 5 and summarized in Figure 6. Actual estimates, based on the game outcome data used in the last section, are shown in Table 3. The estimates are consistent with the effects observed in Figure 4, as is the sizable change in game outcome dynamics after midseason in baseball. The stronger expectations effect in the second half of the season may in part be due to a late mobilization of a national public in anticipation of the World Series. With the large number of local broadcasts siphoning off viewership, networks have for a long time been reluctant to nationally broadcast regular season games before the midseason all-star game.

To eliminate the problem of varying season lengths for cross-sport comparisons, a 48-game season was used for all sports. Toward the same end, 28 team leagues with no divisionalization were used for all sports. To assess the effect of these departures from actuality, the league structure and scheduling constraints (including the 16-game season) of the NFL were used in another set of simulations (for all sports). Every simulation was based on 50 seasons of competition. A season schedule was generated by the computer, and its rows and columns were randomly permuted (within divisionalization constraints, if any) for subsequent seasons to guard against scheduling biases (see Leifer 1990b).

No skill differences were admitted among simulated teams. All teams started the simulation as equals (i.e., past performances of .5), and became differentiated only through the workings of the game outcome equation as the season progressed. For subsequent seasons, however, past season performance was "remembered" with a weight of 4 games (2 for the 16-game simulations) among current season results. Hence, after 4 games into a new season, current season standing is equally weighted with past season results, and after 12 games past season results are three-quarters weighted. By the last game of the season, the past season is forgotten entirely. Obviously, without memory, there could be no relation between performances across seasons, as teams would enter each season as perfect equals and initial differences would occur on a purely random basis.

TABLE 3: Estimates for the Game Outcome Equation Used in the Computer Simulation, Based on Actual 1980 Game Outcomes (1974-80 for Football) in the Four Major League Sports[a]

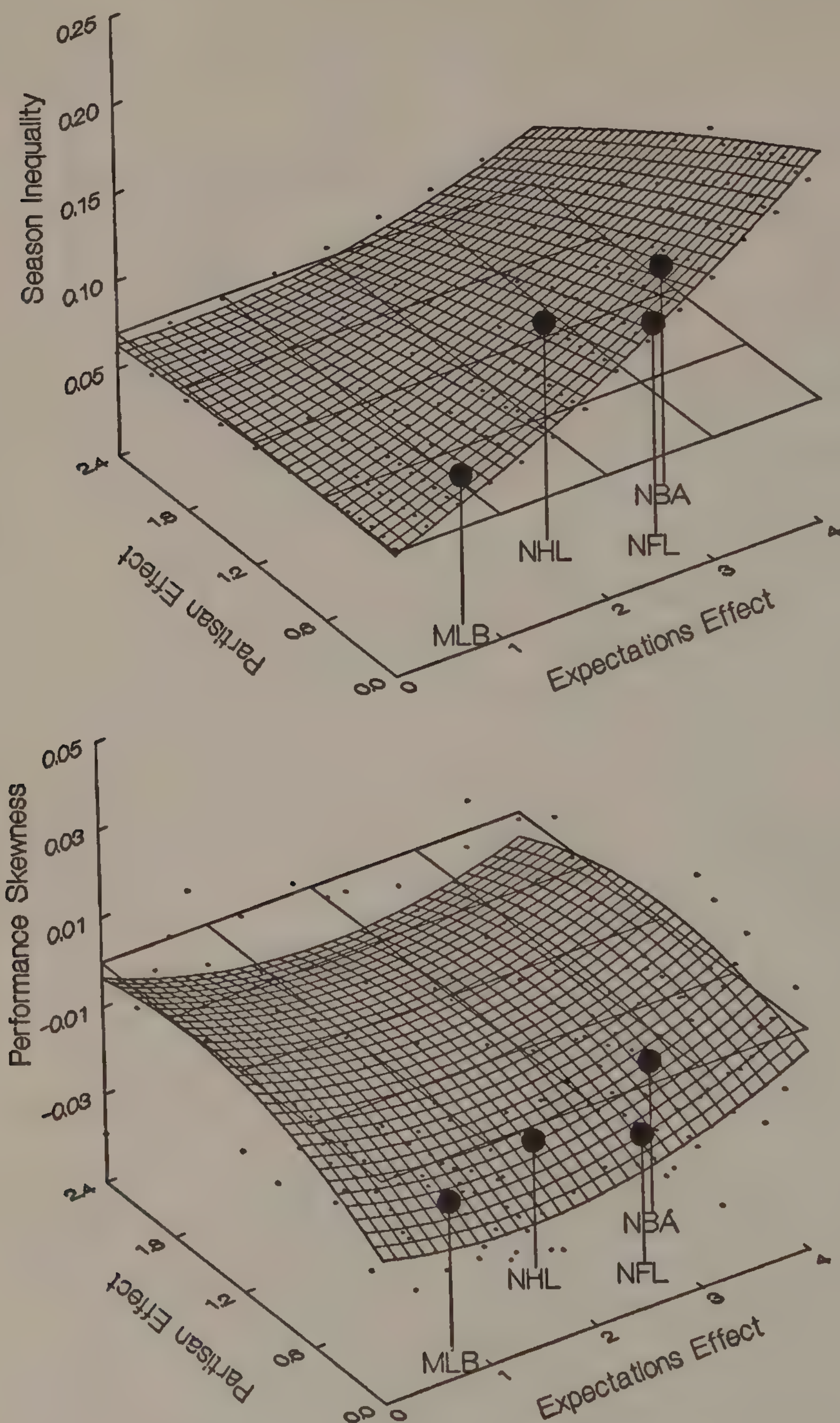
Sport	b_0	b_1	b_2	b_3	Loss Fn
Football	.303	2.437	-.328	.522	661.70
	(.087)	(.266)	(.431)	(.397)	
Baseball (1st half)	.205	.234	.596	-.583	712.80
	(.165)	(.480)	(1.150)	(1.137)	
Baseball (2nd half)	.065	3.216	2.02	-1.902	715.24
	(.220)	(.759)	(1.233)	(1.213)	
Basketball	.641	3.566	-.996	.848	553.06
	(.101)	(.363)	(.537)	(.533)	
Hockey	.525	2.741	.427	-.667	534.68
	(.108)	(.440)	(.643)	(.607)	

[a] Standard errors are in parentheses.

Four measures of performance are monitored, with each measure based on 50 seasons of simulated competition. Season inequality is the average standard deviation of season performances across teams. Performance correlations across seasons gauge the extent to which team performance is stable across seasons (i.e., winners continue winning and losers continue losing). Correlations across both one and four seasons are monitored. Finally, the skewness of the performance distribution is measured by the number of season performances below .5 minus the number above, divided by the maximum possible difference (i.e., one team wins all its games and the rest finish just below .5).

To establish baselines, preliminary simulations varied the levels of the expectations effect ($b_1 \geq 0$) and the constant partisan effect (i.e., $b_0 \geq 0$, with $b_2 = b_3 = 0$). Figure 7 shows the impact of these varying levels on the four performance measures. As the expectations effect increases, season inequality sharply increases over the pure chance level (large grid). The partisan effect operates as a leveler, though substantial results require effects beyond those observed in actual leagues. Similar conclusions can be drawn from looking at inequality across seasons using serial correlations. Correlations spanning four seasons, however, are small even for large expectation effects and small home advantages (this attenuation is largely shaped by memory strength). With no inherent differences between teams (and short memory for emergent differences), national publics can succeed in creating only short-term differences between them. Performance skewness was not systematically related to either the expectation or the partisan effect. Since there is a perfect symmetry between

FIGURE 7A: Simulated Impact of Uniform Partisan and Expectation Effects on Four Performance Measures[a]



[a] Each point used to estimate the fine grid surface is based on a simulation of 50 seasons for 28 competitive equals playing 48 games each. The coarse grid surfaces are the expected results were publics to have no effect on competition.

FIGURE 7B: Simulated Impact of Uniform Partisan and Expectation Effects on Four Performance Measures

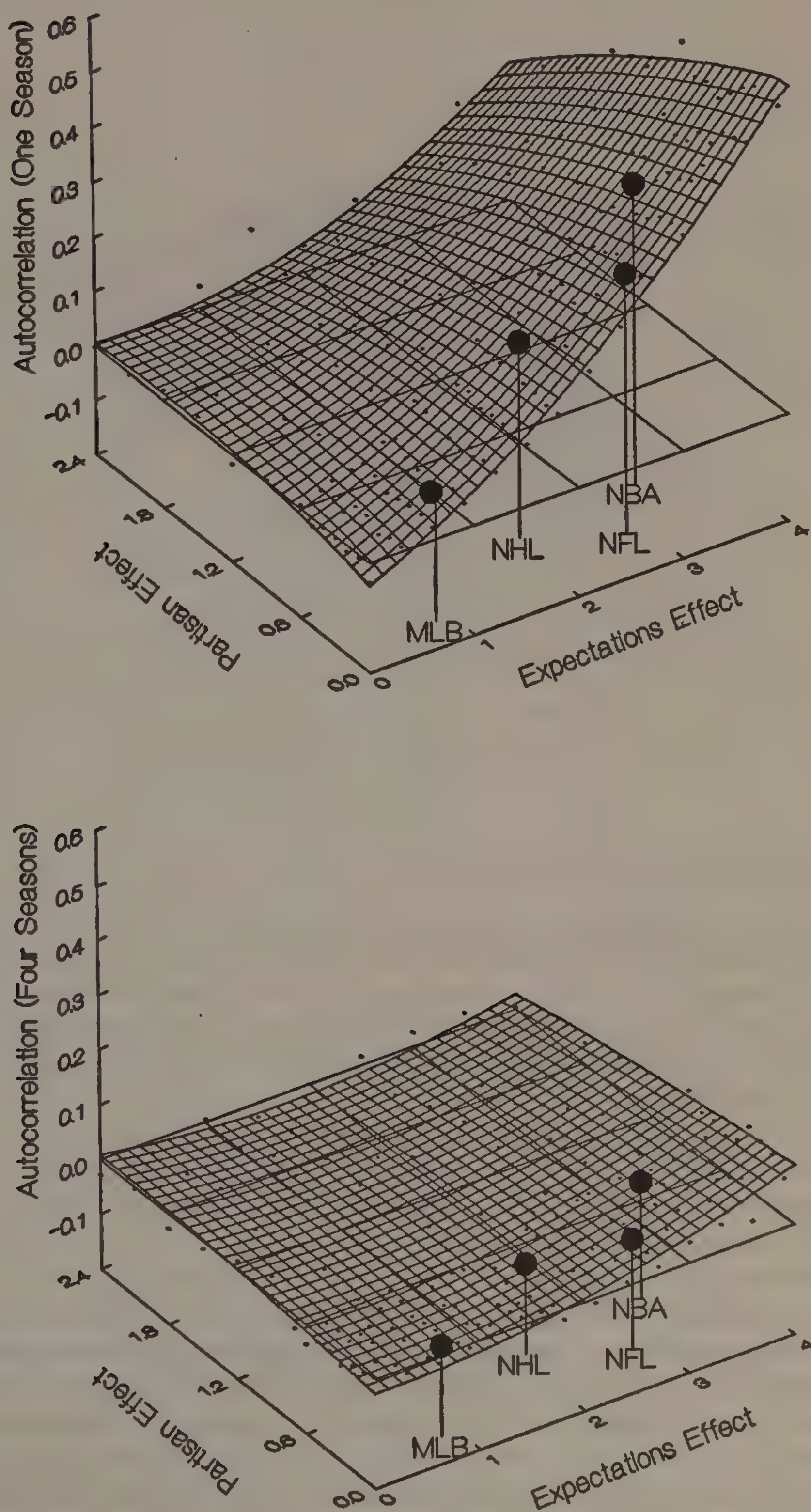
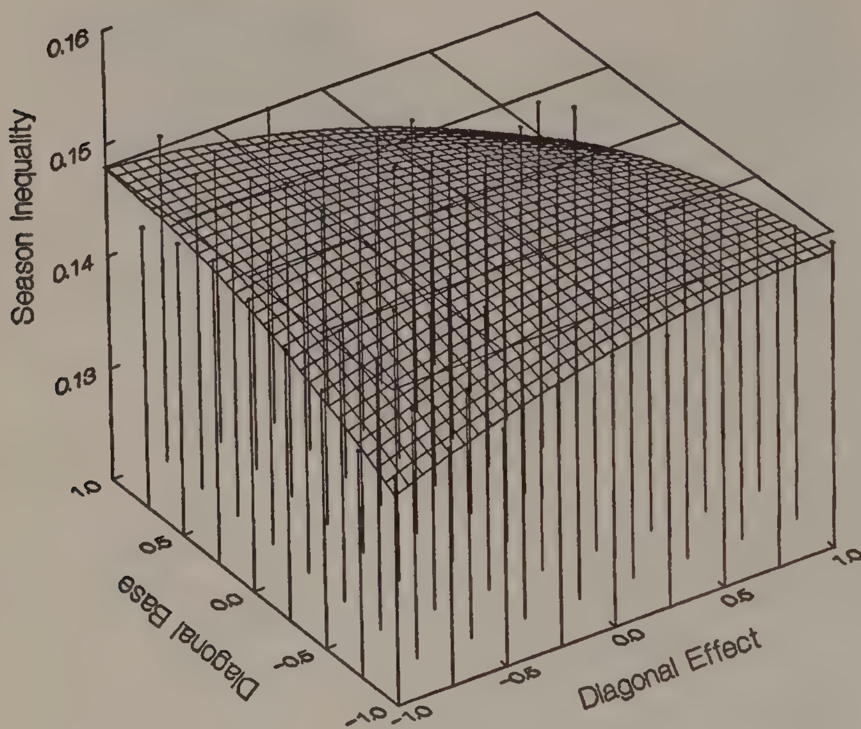


FIGURE 8A: Simulated Impact of Nonuniformities in the Partisan Effect on Performance Measures[a]

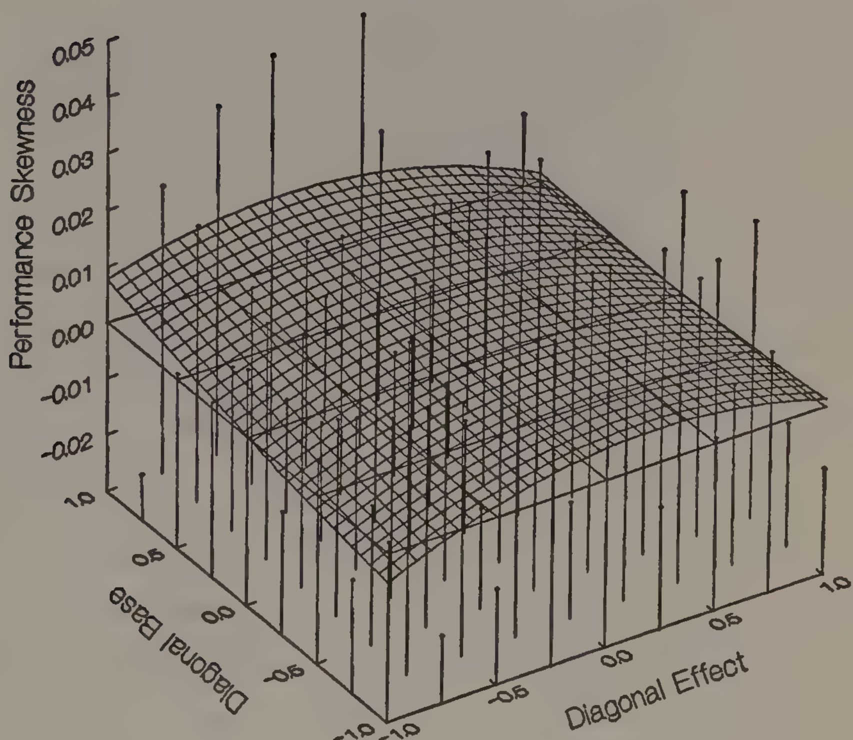


[a] Actual overall partisan and expectation effects from 1974-80 NFL competition are held constant in the simulations (with 28 teams playing 48-game seasons), and partisan effects for similarly performing teams are allowed to vary with performance levels. The diagonal base is the partisan effect for two losing (zero-performance) teams. The diagonal effect is the unit impact of increasing performance levels among similarly performing teams.

high- and low-performing teams in the baseline simulations, there is no reason to expect skewness to be different from zero.

Placing actual leagues in Figure 7 is complicated by the use of a 48-game season. Estimates for the expectation effect (b_1) are sensitive to season length, with longer seasons increasing the effect by enhancing the stability of the performance measure. On the basis of the difference between actual and simulated season length and standard errors, projected (to a 48-game season) estimates for b_1 were 2.750, .788, 3.235, and 1.977 for the sports as ordered in Table 3. A combination of high expectation and low partisan effects generates the most inequality, with the former dominating. Thus if public effects are the

FIGURE 8A: Simulated Impact of Nonuniformities in the Partisan Effect on Performance Measures[a] (Continued)

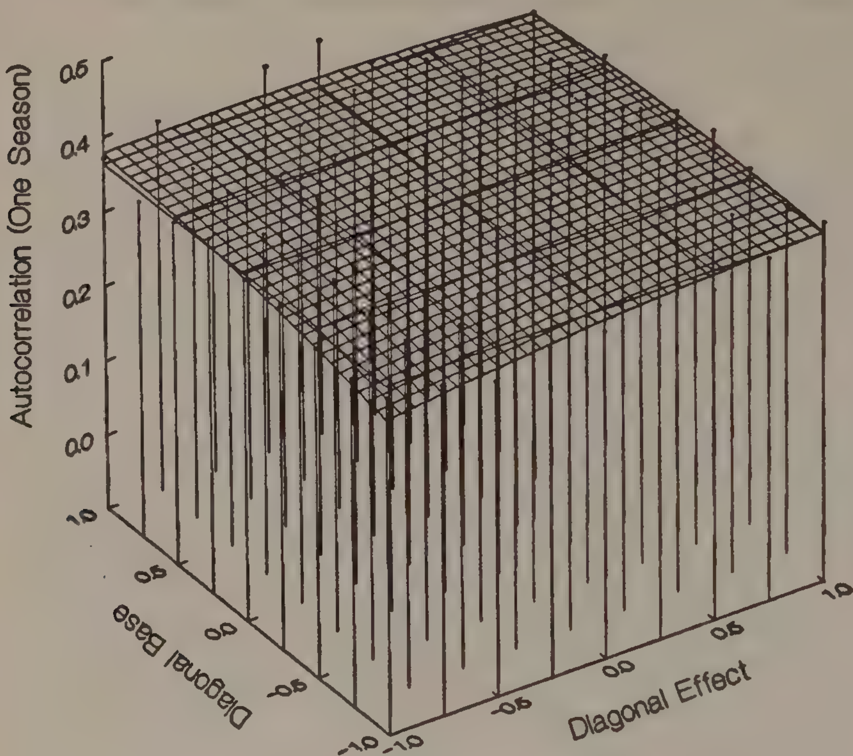


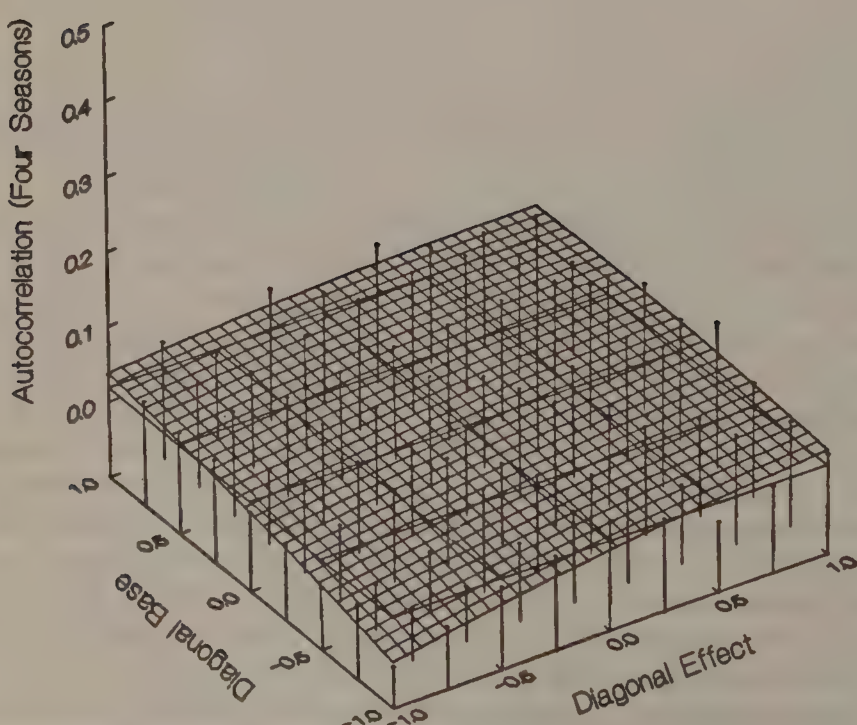
only thing driving performance inequality, football and perhaps basketball should exhibit the most inequality, and hockey and baseball the least.

Nonuniformity in partisan effects were introduced into the diagonal base (b_2) and diagonal effect (b_3) coefficients, using actual estimates for b_0 and b_1 from the four league sports. The results for football (using b_1 = 2.750) are shown in Figure 8. Since a low diagonal base and negative diagonal effect, and vice versa, will affect the overall home advantage level, attention should be directed to the upper-left-to-lower-right diagonal, along which overall home advantage remains constant. It is disappointing that there is not much going on along this diagonal with respect to the inequality measures. At least within the range of actual partisan effect patterns, the nonuniformities are not sufficient to substantially affect long-term performance inequality. Results for the other sports, which are not shown, lead to the same conclusion.

Skewness results were puzzling. Partisan-effect nonuniformities in football and basketball should have produced negative skewness by capping the upper end of performance and allowing spread in the lower end. Positive skewness

FIGURE 8B:


Autocorrelation (One Season)
0.5
0.4
0.3
0.2
0.1
0.0
1.0
0.5
0.0
-0.5
-1.0
Diagonal Base
-1.0
-0.5
0.0
0.5
1.0
Diagonal Effect


Autocorrelation (Four Seasons)
0.5
0.4
0.3
0.2
0.1
0.0
1.0
0.5
0.0
-0.5
-1.0
Diagonal Base
-1.0
-0.5
0.0
0.5
1.0
Diagonal Effect

should be found for baseball and hockey simulations, where the high partisan effect among lower performers should produce the opposite pattern of caps and spreads. It appears, however, that the magnitude of the actual nonuniformities was not sufficient to produce these predicted effects.

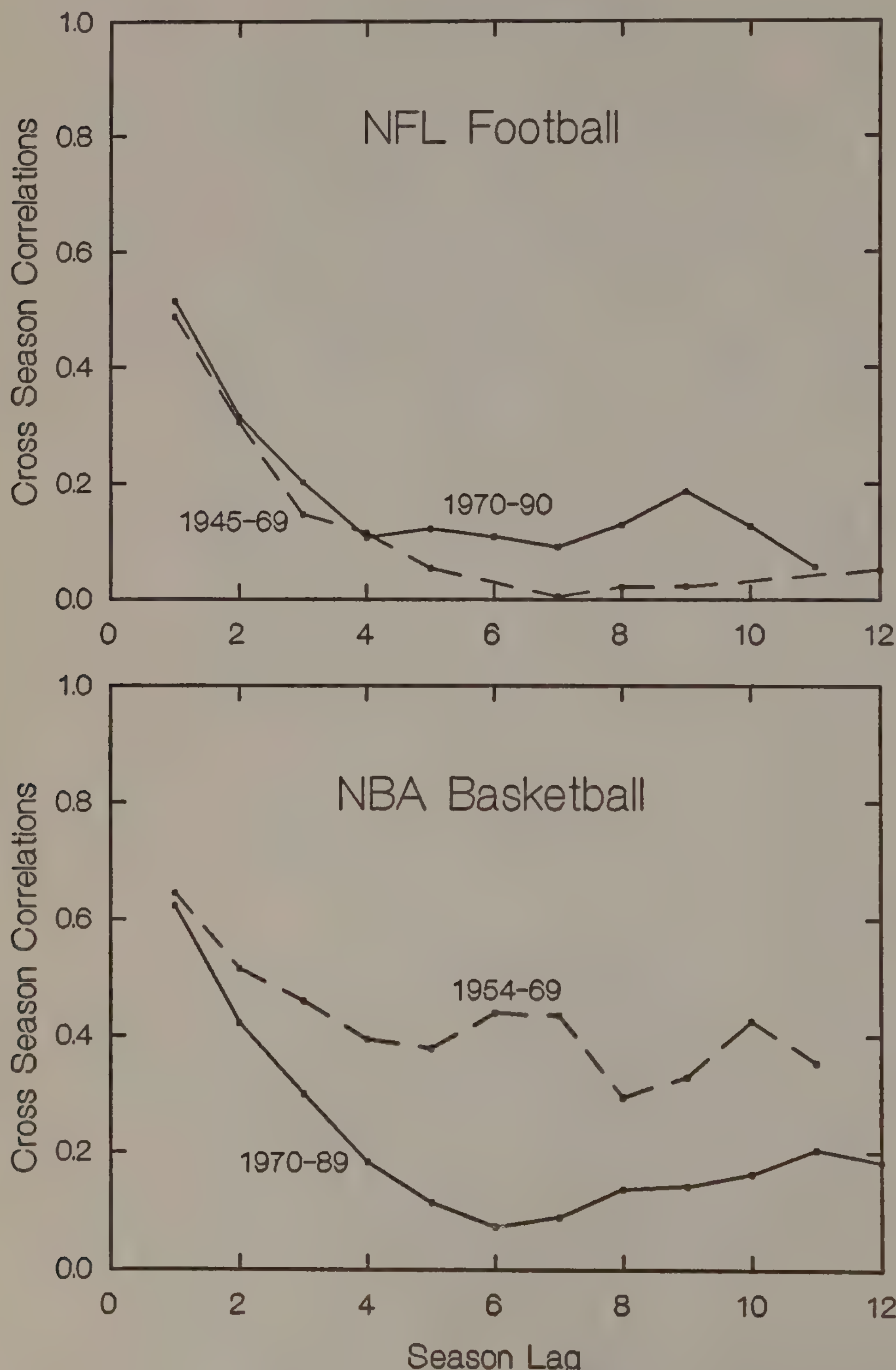
Though the nonuniformities in partisan effects do not appear dramatic enough to affect aggregate performance measures, overall partisan effects and particularly the expectations effect did have substantial effects. Can these factors account for performance measures obtained from the four league sports? Actual performance correlations across up to 12 seasons are graphed in Figure 9. Using the two lower graphs in Figure 7, football results most clearly fit the simulated results for season lags of one and four. Results from simulated 16-game seasons are even more accurate, with cross-season correlations of one and four seasons being .394 and .032 (and a season performance standard deviation of .188, which is within 5% of the actual amount). The other sports, however, should exhibit less inequality than football, since they have either a weaker expectations effect (baseball) or greater home advantage (basketball), or both (hockey). Both baseball and basketball appear to be dramatically moving in this direction considering their much higher inequality in the preceding period, but hockey, with the highest inequality, is moving in the wrong direction.

Missing from the simulation are any factors representing real skill differences between teams. Deviations between the simulated and actual results can be attributed to the persistence of real differences between teams. In this respect, the magnitude of the deviations corresponds closely to the extent to which the sport has entered the modern era of national public cultivation. It is only in this era that true competitive equality has become important, as it induces fans to search broader and deeper for winners than past local loyalties would even allow. In the prototypically modern sport of football, nearly all persisting inequality can be accounted for by the mix of support and expectations that publics impose on sporting competition. The success of the NFL in cultivating national publics has been facilitated by the elimination of real differences between teams. Real differences keep hockey from moving beyond the local loyalties it still depends on.

Performance measures for individual seasons (1970-90) are shown in Figure 10. Season inequality is closely tied to season length in the top panel. Dotted lines provide the theoretically expected inequality were every game to be decided by a coin flip (i.e., competitive equality). Relative to this baseline, results conform to those for cross-season inequality except for hockey, which exhibits less inequality than basketball. Though this may be due to the difficulty of scoring in hockey and the admission of ties (which are numerous), the fact that season inequality is decreasing (A < B) over the period should be noted. On this note, the season-to-season correlations (aggregated in Figure 9) in the bottom panel show the same pattern of decreasing inequality in the 1980s for hockey.

Again, skewness results (middle panel) prove hard to interpret, particularly in the absence of clear simulation results. The 1980 season, used to estimate the game outcome equation, is clearly an outlier for basketball and somewhat for hockey. Basketball shows a clear trend toward negative skew, implying a performance cap that makes for close championship races combined with a

FIGURE 9A: Actual Correlations between Performances in Seasons *t* and *t* lag across All Teams and Seasons within Indicated Periods for Each Major League Sport[a]



[a] Lag periods that bridge a team move were excluded.

FIGURE 9B: Actual Correlations between Performances in Seasons *t* and *t* lag across All Teams and Seasons within Indicated Periods for Each Major League Sport[a]

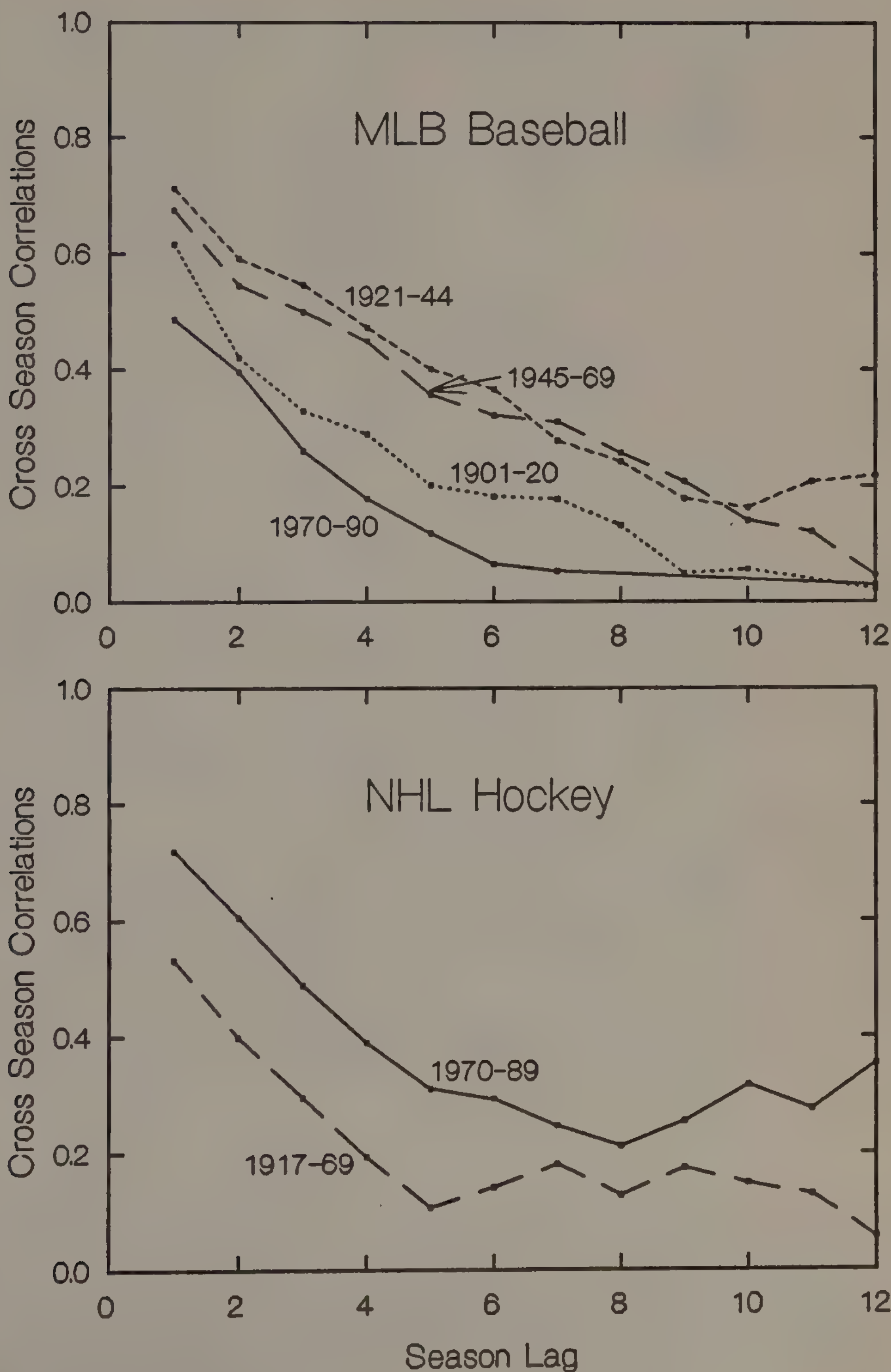
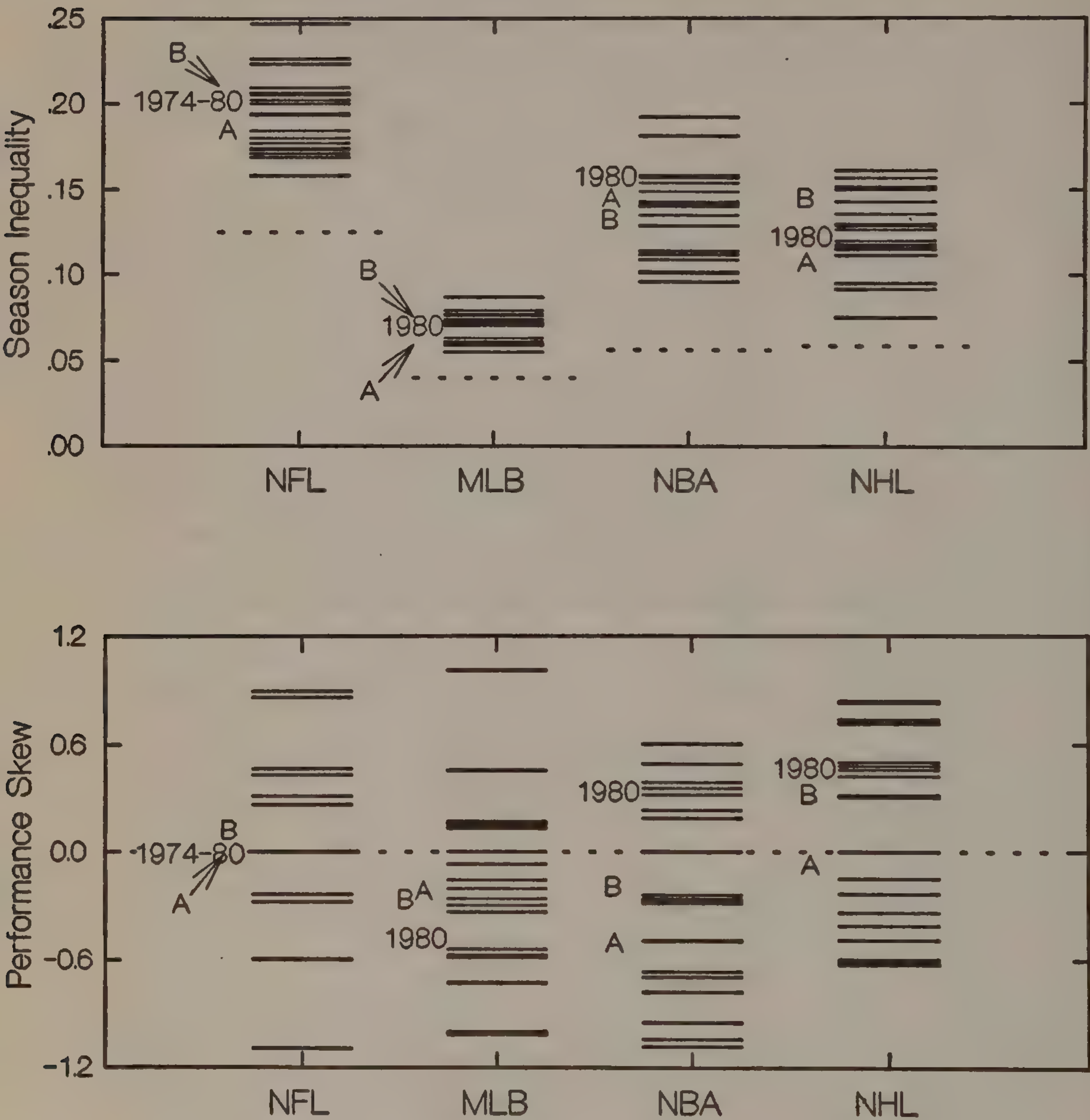
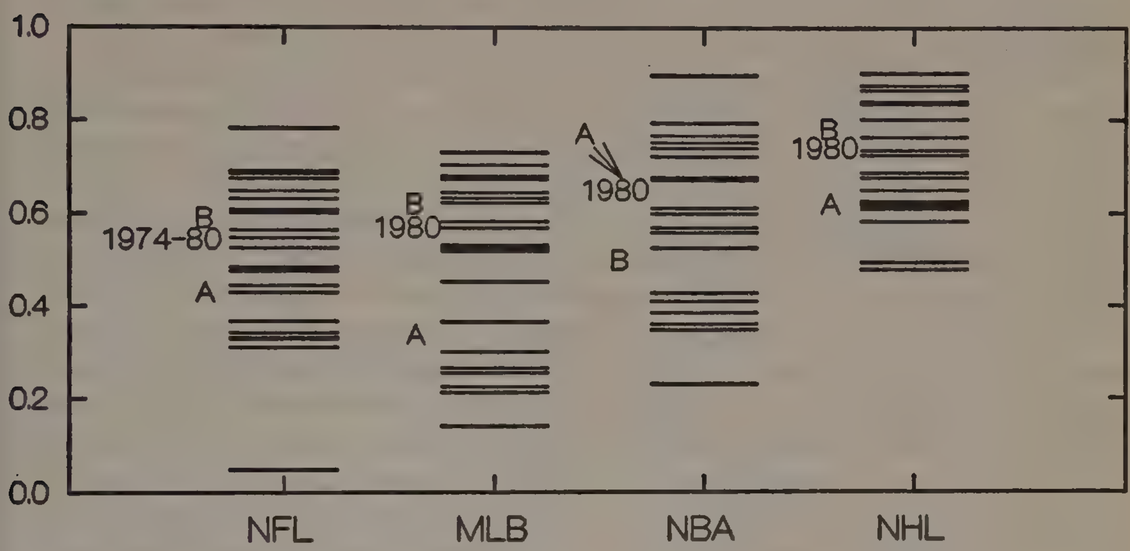


FIGURE 10: Actual Performance Measures for 1970-90 Seasons in the Four League Sports[a]



[a] The 1980 season (1974-80 in football) provided estimates for the partisan and expectations effects used in the simulation and is marked along with averages for 1970-79 (B) and 1981-90 (A). Measures in the three panels are those used for the simulation results, except that they are based on individual seasons rather than aggregated data for 50 seasons.

FIGURE 10: Actual Performance Measures for 1970-90 Seasons in the Four League Sports[a] (Continued)



large spread at the bottom of the performance distribution. But this pattern is only weakly found in football, where it was most expected, and the expected converse pattern is not to be found in baseball and hockey. Hence the actual and simulated results are consistent, but both fail to confirm the impact of publics that was predicted before the simulations were run.

In sum, the expectation effect dominated the simulation results, but it could be muted somewhat by a strong overall partisan effect. No clear further impact could be found for nonuniformities in the partisan effect. Performance predictions based solely on public effects were most accurate for football, the sport that has most successfully cultivated national publics. Deviations for the other sports, in the form of real performance differences between teams, varied directly with the progress the sport had made in cultivating national publics. Hence the successful pursuit of competitive equality can be tied to entry into the modern era of national television publics. But the competitive equality remains beneath the surface as national publics induce performance inequality by casting support for winners and withdrawing support from losers.

Discussion and Conclusion

Social support is something that is given and received within an interaction. Both the giving and receiving have consequences not only for those directly involved, but also for those excluded. Because social support involves a taking of sides, it is intrinsically constrained in any given setting. What is given to one side is withheld from another side. The widely observed fact that social support is linked to a diverse array of positive outcomes can lead to misleading recommendations if this constraint is not taken into account. Social support is

sought out and proffered amidst contentions, and every public instance of it marks a defeat as well as a victory.

That social support is received at all is inherently problematic. Even within families, support can be the exception and not the rule, and it can trigger tensions where it is too readily activated, as it is with ailing children or aging parents (Abel 1989; Cooney & Uhlenberg 1992; Hogan, Eggebeen & Clogg 1993). It was necessary, in this article, to look at how competition was organized by major leagues to anticipate where public support for teams might appear and what might be its latitude for affecting outcomes. The central finding of this study is that social support has a perverse effect. It works against the basis on which it is organized. Where major leagues encourage large-city dominance, partisan local publics function as a performance leveler. Conversely, where major leagues pursue competitive equality, expectations of national publics function to induce performance inequality. The more effectively a major league organizes competition for one or the other public, the more effect the publics will have on performance outcomes. No sport has been more effective in cultivating national publics than football, and here we found that the expectations effect could account for virtually all the actual performance inequality. In hockey, still almost entirely dependent on local publics, real inequality persists that a strong partisan effect (especially for low performers) only serves to dampen somewhat.

Organizers and publics are, in effect, locked in an oddly reinforcing antagonism. In sports dependent on local publics, disadvantaged smaller-city teams depend on loyal partisan support to survive in both economic and performance terms. Partisan support helps smaller-city teams win at home and hence directly implicates the local public in performance outcomes. Yet the sizable support that a weak small-city team can maintain only encourages leagues to pursue large-city dominance, as smaller cities would not generate more support by winning than large cities would lose if they failed to win. As long as the attachment between teams and cities remains alive, major league encouragement of large-city dominance merely fuels the need for smaller cities to support their teams, to economically survive and sometimes win.

A dependence on national publics reverses these roles. The more effectively major leagues pursue competitive equality, the more broadly and energetically national publics search for winners to follow. While these publics may appear eager to find real heroes in scouring league competition, the emergence of a truly superior team would dull their interest in the game. Major leagues are typically credited for the great teams they produce, but their policies are directed at spreading talent across teams and not concentrating it within a single one (Leifer 1995). By organizing to reduce real differences between teams, the NFL has allowed publics to freely construct (and deconstruct) the heroes they are looking for. Real performance differences between teams persist in the other sports, particularly hockey, which may only reinforce local dependencies and hinder the emergence of national publics.[12]

Few settings are as tightly organized as major league sports, and hence we cannot expect other types of publics to play as great a role as they do in sports. Yet there are many settings where the dynamics found above have appeared in cruder form. Rosenbaum's (1976, 1984) studies of tracking in high schools and

career mobility in corporations offer intriguing examples. The more care that is taken in either case to place initial recruits on an equal footing, the more energy is spent in finding differences between them once school or jobs have begun. The publics may be partisan on-site teachers or department supervisors, or they may be bearers of universalistic standards lodged in central administration as counselors or personnel experts. This should make a difference in the inequalities that emerge, if anything is to be learned from the present study (Berlew & Hall 1966).

Even where the "design" of a system is purely evolutionary, the complex workings of social support can play a crucial role in shaping inequalities. Cohen (1981), in a study of the Creoles in Sierra Leone, found ambiguities in the definition of "family" that left openings for both partisan and expectation effects. A "cult of the dead" tied the biological offspring of a deceased person together in that property could not be divided among surviving relatives. This became the basis by which lower-status relatives could make claims on higher-status ones. Otherwise, Creoles were endlessly trying to establish "cousinhood" by gaining inclusion into the ceremonials of well-positioned persons (and avoiding lower-status claimants to cousinhood with them). Guest lists throughout Creoledom were closely monitored, for who is in and out in the scramble for advantageous association, with "family" here defined by who showed up at one's ceremonials with gifts of cake and drink in hand. Cohen recognized that the cult of the dead served as a leveler that helped prevent permanent class-based rifts from developing, while the cousinhood dynamics induced pronounced inequalities within Creoledom.

In all the above cases, a random sampling of individuals would confirm that those receiving more social support experienced more positive outcomes. But this finding ignores both the problem of how social support comes to be generated and the perverse effects it has where it appears. In a similar vein, Boudon (1982) has argued that the fact that education is associated with status attainment across individuals does not mean that increasing everyone's education level will increase everyone's status attainment. The workings of social support can be even more perverse in moving beyond the individual level because of the immense significance that social support can have for those who see it being directed toward someone else, with whom they may be contending for status.

Notes

1. Efforts have been made to explain the home advantage in terms that do not directly involve the support of local publics. Travel fatigue among visiting teams is an obvious possibility, yet it fails on close examination. With 162 games played in less than six months, travel fatigue should be greatest in baseball, yet here the home advantage is weakest. Even within (minor league) baseball, Courneya and Carron (1991) found no consistent evidence that travel is a major contributing factor to the home advantage in a detailed statistical examination of direct indicators of travel fatigue (i.e., length of visitor's road trip, length of home stand, series game number, and season game number). The same authors (1990) eliminate a second possibility based on the notion that rules such as the home team's right to bat last in baseball confer an advantage. A learning factor, where the home advantage is traced to the home team's superior knowledge of the local facilities and conditions, has fared little better (Pollard 1986). Baseball

stadiums are among the least standardized of all playing arenas, with different turfs, playing field sizes, wall heights, and local weather conditions.

2. Publics can put pressure on competitors to assume winner and loser roles within a game or pennant race that might not emerge otherwise. Timing is crucial, as stepping into a winner role before it is conferred is strategically unsound. Clausewitz ([1826] 1976) analyzed the weakness of assuming the aggressor role in the absence of superior strength. Before the aggressor-defender roles that lead to winning and losing are assumed, teams must engage in probes and subtle ploys that can themselves conceal opportunities for clear action. Publics can play a decisive part in disrupting the stable equilibrium by latching onto apparent inequalities and convincing both sides of their reality. If partisanship or expectations encouraged one home team into the actions of a winner without simultaneously casting the other team into a loser's role, the actions would backfire. Under pressure from publics, subtle actions give way to actions appropriate to a winner and loser. Without publics, there may be no reliable basis for claiming an advantageous role. For a more thorough discussion of these issues, see Leifer (1988, 1991).

3. Ideally, the size of a stadium would be based on local tendencies to support a home team. If the size of local stadiums stood in direct relation to the size of crowds that would show up for the most appealing games, then actual crowd density would be an indicator of relative appeal. Crowd densities of, say, .6 would mean the same everywhere — 60% of potential support is being tapped by the particular contestants (ignoring consequences of the particular day and time they meet). Unfortunately, the politics of stadium construction do not always yield close fits between stadium size and potential support. Overall, large stadiums prove harder to fill regardless of where they are located or the teams that play in them. Hence in trying to account for crowd density, stadium size had to be (re)introduced as a control variable.

4. In major league baseball, there were only 17 sellouts of which the Kansas City Royals (the pennant winner) accounted for 7 and the Texas Rangers for 4. Though sporadic with regard to home-away performance balance, all sellouts were elicited by games where average performance levels were high.

5. When the National League and American League are analyzed separately, the rise is entirely confined to the NL (.130 vs. .000, respectively), making the phenomenon league-specific. Since there is no reason to expect league differences here, any explanation would be hazardous without more data.

6. The dependent variable is actually game outcomes (code 0, .5, and 1 for losses, ties, and wins). HWC are predicted values. Logistic regression was used in the simulation to ensure that these values do not stray outside the 0,1 range. OLS estimates are used here because they yield similar results and are easier to interpret in impact terms. Home advantage can be calculated directly from the HWCs for performance combinations above and below the equal performance diagonal, and hence factors that affect the latter necessarily affect the former. The dependence of home advantage on outcomes in both home and away locations, however, makes it difficult to analyze directly.

7. During 1980 basketball's Seattle Supersonics played in the Kingdome (where Seattle's baseball team played), which was almost twice as large as the second-largest arena (40,248 versus 22,422). I used the second-largest to scale capacities for basketball and ran analyses involving stadium capacity both with and without Seattle. The text notes where this exclusion made a difference.

8. Within each sport, performance as a measure of skill would imply greater predictive ability in the second half of the season than the first. Yet only in baseball was this the case, with R^2 rising from .000 to .024. In the other three sports, R^2 declined by 15% to 25% in the second half of the season.

9. The use of performance differences (as opposed to using both home and away performance levels), for example, does not allow HWCs to vary along the same performance diagonal. Crowd density and stadium capacity were introduced to account for some of the observed nonuniformities in home advantage. Technically, their influence was not confined to the residuals, since coefficients for performance differences were reestimated with the introduction

of these variables. The coefficients changed very little, however, outside of a slight increase (10%) in the baseball coefficient.

10. When the impact of stadium capacity and crowd density was allowed to vary across performance regions, the impact of performance differences entirely disappeared in baseball. This was not the case in any of the other sports, where performance differences continued to be influential (becoming even more so in hockey and not changing at all in football). While separate intercepts for each performance region could absorb regional variation in game outcomes, the fact that this undermined the impact of performance differences only in baseball is quite puzzling.

11. To eliminate any abrupt discontinuities between equals and unequals, an interpolation between the two levels was used for contenders whose performance levels differed by .1 to .2.

12. Baseball is in a precarious position. Its early attachment to cities was exploited by increasing the number of games per season, to generate more gate receipts. But the large number of games became a liability for cultivating national publics, which requires limiting the number of games broadcast. The recent drift toward greater performance equality is a positive development only if a national public is mobilized and succeeds in inducing performance inequality.

References

Abel, Emily K. 1989. "The Ambiguities of Social Support: Adult Daughters Caring for Frail Elderly Parents." *Journal of Aging Studies* 3:211-30.

Allport, Frank H. 1920. "The Influence of the Group upon Association and Thought." *Journal of Experimental Psychology* 3:159-82.

Andrew, John A., Jr. 1984. "An Analysis of Home and Away Game Performance of Atlantic Coast Conference Basketball Teams." Ph.D. diss., University of North Carolina, Chapel Hill.

Bainbridge, William S. 1990. "Explaining the Church Member Rate." *Social Forces* 68:1287-96.

Berlew, David E., and Douglas T. Hall. 1966. "The Socialization of Managers: Effects of Expectations on Performance." *Administrative Science Quarterly* 11:207-23.

Boudon, Raymond. 1982. *The Unintended Consequences of Social Action.* St. Martin's Press.

Chase, Ivan D. 1984. "Social Process and Hierarchy Formation in Small Groups: A Comparative Perspective." In *Social Hierarchies: Essays Toward a Sociophysiological Perspective,* edited by Paula R. Barchas. Greenwood Press.

Clausewitz, Carl von. [1826] 1976. *On War.* Princeton University Press.

Cohen, Abner. 1981. *The Politics of Elite Culture: Explorations in the Dramaturgy of Power in a Modern African Society.* University of California Press.

Cooney, Teresa, and Peter Uhlenberg. 1992. "Support from Parents over the Life Course: The Adult Child's Perspective." *Social Forces* 71:63-84.

Cornell, L.L. 1992. "Intergenerational Relationships, Social Support, and Mortality." *Social Forces* 71:53-62.

Cottrell, N.B. 1972. "Social Facilitation." *Experimental Social Psychology,* edited by C.G. McClintock. Holt, Rinehart & Winston.

Courneya, Kerry S., and Albert V. Carron. 1990. "Batting First versus Last: Implications for the Home Advantage." *Journal of Sport and Exercise Psychology* 12:312-16.

_____. 1991. "Effects of Travel and Length of Home Stand/Road Trip on the Home Advantage." *Journal of Sport and Exercise Psychology* 13:42-49.

_____. 1992. "The Home Advantage in Sport Competitions: A Literature Review." *Journal of Sport and Exercise Psychology* 14:13-27.

Dekeseredy, Walter S. 1990. "Male Peer Support and Women Abuse: The Current State of Knowledge." *Sociological Focus* 23:129-39.

Dowie, J. 1982. "Why Spain Should Win the World Cup." *New Scientist* 94:693-95.

Edwards, John. 1979. "The Home Field Advantage." In *Sports, Games and Play: Social and Psychological Viewpoints*, edited by J.H. Goldstein. Erlbaum.

Greer, Donald L. 1983. "Spectator Booing and the Home Advantage: A Study of Social Influence in the Basketball Arena." *Social Psychology Quarterly* 46:252-61.

Haines, Valerie A., and Jeanne Hurlbert. 1992. "Network Range and Health." *Journal of Health and Social Behavior* 33:254-66.

Hogan, Dennis, David Eggebeen, and Clifford Clogg. 1993. "The Structure of Intergenerational Exchange in American Families." *American Journal of Sociology* 93:1428-58.

Hougland, James G., Dwight Billings, and James Wood. 1990. "The Instability of Support for Television Evangelists: Public Reactions during a Period of Embarrassment." *Review of Religious Research* 32:56-64.

House, James S., Debra Umberson, and Karl R. Landis. 1988. "Structure and Process of Social Support." *Annual Review of Sociology* 14:293-318.

Ibarra, Herminia. 1992. "Homophily and Differential Returns: Sex Differences in Network Structure and Access in an Advertising Firm." *Administrative Science Quarterly* 37:422-47.

Klatell, David A., and Norman Marcus. 1988. *Sports for Sale: Television, Money, and the Fans*. Oxford University Press.

Leifer, Eric M. 1988. "Interaction Preludes to Role Setting: Exploratory Local Action." *American Sociological Review* 53:865-78.

____. 1990a. "Inequality among Equals: Embedding Market and Authority in League Sports." *American Journal of Sociology* 96:655-83.

____. 1990b. "Enacting Networks: The Feasibility of Fairness." *Social Networks* 12:1-25.

____. 1991. *Actors as Observers: A Theory of Skill in Social Relationships*. Garland.

____. 1995. *Making the Majors: The Rise of Major Leagues and the Publics They Created*. Harvard University Press.

McGuire, E.J., Kerry S. Courneya, W. Neil Widmeyer, and Albert V. Carron. 1992. "Aggression as a Potential Mediator of the Home Advantage in Professional Ice Hockey." *Journal of Sport and Exercise Psychology* 14:148-58.

Merton, Robert K. 1957. *Social Theory and Social Structure*. Free Press.

Mizruchi, Mark S. 1985. "Local Sports Teams and Celebration of Community: A Comparative Analysis of the Home Advantage." *Sociological Quarterly* 26:507-18.

Paulus, P.B., and W.L. Cornelius. 1974. "An Analysis of Gymnastic Performance under Conditions of Practice and Spectator Observation." *Research Quarterly* 45:56-63.

Pescosolido, Bernice A. 1990. "The Social Context of Religious Integration and Suicide." *Sociological Quarterly* 31:337-57.

Podolny, Joel. 1992. "The Matthew Effect and the Constraints of Status: The Investment Banking Industry." Manuscript, Stanford Business School, Palo Alto, California.

Pollard, R. 1986. "Home Advantage in Soccer: A Retrospective Analysis." *Journal of Sport Sciences* 4:237-48.

Quirk, James, and Rodney D. Fort. 1992. *Pay Dirt: The Business of Professional Team Sports*. Princeton University Press.

Rader, Benjamin G. 1984. *In Its Own Image: How Television Has Transformed Sports*. Free Press.

Rosenbaum, James E. 1976. *Making Inequality*. Academic Press.

____. 1984. *Career Mobility in a Corporate Hierarchy*. Academic Press.

Schwartz, Barry, and S.F. Barsky. 1977. "The Home Advantage." *Social Forces* 55:641-61.

Stevens, Ellen S. 1992. "Reciprocity in Social Support: An Advantage for the Aging Family." *Families in Society* 73:533-41.

Thirer, J., and M.S. Rampey. 1979. "Effects of Abusive Spectator's Behavior on Performance: Visiting Intercollegiate Basketball Teams." *Perceptual and Motor Skills* 48:1047-53.

Travis, L.E. 1925. "The Effect of a Small Audience upon Eye-Hand Coordination." *Journal of Abnormal and Social Psychology* 20:142-46.

White, Lynn K., and Agnes Riedman. 1992. "Ties among Adult Siblings." *Social Forces* 71:85-102.

Whyte, William F. 1946. *Street Corner Society*. Free Press.

Zajonc, R.B., and S.M. Sales. 1966. "Social Facilitation of Dominant and Subordinate Response." *Journal of Experimental Social Psychology* 2:160-68.

Zeller, R.A., and T. Jurkovac. 1988. "Dome-inating the Game." *Psychology Today* Oct. pp. 19-22.

General Social Survey Can Now Be Searched Electronically

The *General Social Survey (GSS)* is now part of the electronic Public Opinion Item Index maintained by the Institute for Research in Social Science *(IRSS)* at the University of North Carolina at Chapel Hill. A national survey conducted yearly by the National Opinion Research Center *(NORC)* since 1972, the *GSS* repeats many questions regularly, allowing researchers to track changes in opinion on a variety of issues. As part of the *IRSS* Index, *GSS* questions and their marginal frequencies can be examined by researchers from around the world via the INTERNET or by phone.

Who Said There's No Free Lunch?

The *IRSS* Index allows any researcher with a modem-equipped PC or connection to INTERNET to view questions and the frequency distributions of the responses for the *GSS*, as well as a variety of state and national polls, including the Louis Harris Polls. The service is free (except for the long-distance charge incurred by modem users). Using the Index, researchers can search electronically for specific poll questions by using combinations of words which occur in the questions. The Index provides quick and easy access for users who want to find previous question wordings in order to develop questions for their own studies, as well as for researchers interested in obtaining frequencies or in locating particular variables for statistical analysis.

More Free Stuff

For detailed information on accessing the item index, performing searches using Boolean search techniques, and printing or using electronic mail to retrieve results, users can obtain the free booklet "A Guide to Searching the IRSS Data Archives" from *IRSS*. To obtain the booklet by mail, send an email message containing your name, mailing address and a request for the "Guide to Searching" to Bonita Samuels at *BSAMUELS.IRSS@MHS.UNC.EDU*. If you wish a copy sent via electronic mail, please mention this in your request and include a complete email address.

How to Access *FREE* IRSS Data Services via the INTERNET

For those who wish to log on using INTERNET, *UNCVM1.OIT.UNC.EDU* is our INTERNET address. After connecting, you will be asked to log on. At the prompt enter a userid "irss10", "irss11" or "irss12". Press return, then enter the password "IRSS". Press return ONCE. You will see "VM read" in the lower right-hand corner of the screen. Press return ONCE and the message should change to "running". (If you get a message that the logon is in use, try another one.) Note: If at any point you see "MORE" or "HOLDING" in the lower right-hand corner of the screen, press the clear key ("Esc ." for VT100 emulation). Next, you will see a screen announcing IRSS's Data Services. From that point on, procedures are self-explanatory.

Any Questions?

For more information about the Public Opinion Item Index or how to get data including the *GSS*, contact David Sheaves at (919) 966-3348, by email at *UIRDSS@UNCVM1.BITNET*, or by fax at (919) 962-4777.

Negotiated Exchanges in Social Networks*

MICHAEL J. LOVAGLIA, *University of Iowa*
JOHN SKVORETZ, *University of South Carolina*
DAVID WILLER, *University of South Carolina*
BARRY MARKOVSKY, *University of Iowa*

Abstract

Network exchange theory predicts relative profits from negotiations among actors in social exchange networks (Markovsky et al. 1993; Markovsky, Willer & Patton 1988). Here we extend the theory to allow exact predictions, rather than merely ordinal, for actors' exchange profits. This is accomplished by integrating two important factors. First, a resistance model predicts bilateral negotiation outcomes within a given set of network constraints. It does so by weighing actors' interests in gaining the best possible exchanges against their desires to avoid the worst. Second, the resistance model predictions are modified by actors' profit expectations. In particular, we incorporate two factors that affect such expectations, both common features of ongoing exchange relations: the number of other actors to whom one is directly connected in the network, and the likelihood of one's completing exchanges with them. We derive hypotheses from the theory and test them in two very different experimental settings. We find that the theory's predictions are more accurate than those of previous versions of the theory and those of five alternative theories.

Social exchange theory grew from the application of the economic theory of exchange to social relationships. Sociology focuses on a problematic area for economic theory: the exchange of valued objects in relatively small groups, where actors seek to settle on one optimal outcome out of a range of possibilities. How can we predict that outcome? And how are such outcomes affected by social structure? Homans (1958, 1974) suggested that principles of behaviorist psychology would help to answer these questions. Blau's (1964) approach used rational choice and utility theory. Theoretical work on the problem since then has largely developed from one or the other perspective, sometimes combining the two.

* *We gratefully acknowledge the National Science Foundation for supporting this research (Grants SES #88-08289 and SES #90-22935 to Markovsky, SES #90-10888 and #92-23799 to Skvoretz and Willer, and SES 92-23688 to Markovsky and Lovaglia).* *Direct correspondence to Michael J. Lovaglia, Department of Sociology, University of Iowa, Iowa City, IA 52242. E-mail: Michael-Lovaglia@uiowa.edu.*

 Social Forces, September 1995, 74(1):123-55

Both Homans and Blau illustrated their ideas in the context of dyadic relationships. In work settings, for instance, an exchange may entail conferring some reward in return for a costly act. Giving prestige in return for expert help is a common example. Some of these illustrations were quite ingenious. Their very ingenuity, however, convinced critics that the social exchange perspective was tautological and scientifically vacuous. It seemed that any outcome could be explained by a judicious identification of the costs and rewards. This criticism was difficult to overcome as long as the theoretical context remained the isolated dyad. The key to theoretical advance in social exchange was to focus on the embeddedness of dyadic relations in broader contexts. The conceptual horizon has since expanded to incorporate broader relational structures. As a result, social networks have become the focus for rigorous tests of developing theories. These developments were led by Emerson's (1962) seminal work on power-dependence theory. He was the first to suggest specific ways to extend a model of dyadic exchange to larger networks of exchange relations (Emerson 1972). Emerson and colleagues (Cook et al. 1983) later introduced the concept of vulnerability, a measure that predicted which positions in a network structure had power. Vulnerability was based on the idea that some network positions are more important than others in determining the flow of resources through a network. If by removing itself from an exchange a position could reduce the total resources available in a network, then that position had power. The amount of power depends on the disruption in resource flow.

The pace of research on structural power in social exchange networks quickened soon thereafter, spurred by Willer's (1986) critique of the 1983 vulnerability model and a reply by the power-dependence group (Cook, Gillmore & Yamagishi 1986). It was another two years before an alternative formulation was developed that offered better predictions — the network exchange theory of Markovsky, Willer, and Patton (1988). To predict relative power levels for positions in exchange networks, it provided a graph-theoretical power index (GPI) based on a network path-counting algorithm. The theory challenged some basic assumptions of power-dependence theory and withstood critical tests that corroborated GPI predictions and falsified those from the vulnerability model of power-dependence theory.

In 1990 comments and responses pointed out limitations of the GPI (Markovsky, Willer & Patton 1990; Yamagishi & Cook 1990). Soon thereafter, Markovsky (1992) introduced a further refinement, just as power-dependence researchers replaced vulnerability with a completely new algorithm (Cook & Yamagishi 1992). At the same time, three new theories were introduced: Friedkin's (1992) expected value theory, Bienenstock and Bonacich's (1992) application of the "core" from game theory, and Skvoretz and Fararo's (1992) application of Coleman's (1990) rational exchange model. In the next year, Markovsky et al. (1993) identified a new class of structural dynamics and additional refinements in the network exchange theory. Most recently, Skvoretz and Willer (1993) tested the first ratio scale predictions from four theories on a variety of networks. Now, it seems, theoretical developments that once required half a decade occur within a year's time.

Central to theoretical development today is a class of networks in which subtle power differences occur. This phenomenon is known as weak power. In these weak power networks, some positions may have advantages over others in acquiring resources through exchange. However, unlike the advantages in strong power networks, advantages in weak power networks are not progressive. Over a series of exchanges, a strong power advantage eventually results in one exchange partner receiving nearly all available resources. Weak power is limited in range and magnitude. This necessitates theoretical refinement because adequate assessment of power differences between positions in weak power structures requires more precise predictions of exchange rates at equilibrium.

Network exchange theory (Markovsky, Willer & Patton 1988) and its weak power extension (Markovsky et al. 1993) generate ordinal predictions for profits accruing from exchanges among negotiating actors in social networks. The theory is supported primarily by data from experiments in which actors have full information about the shape of the network and know the offers and agreements of all other actors (Skvoretz & Willer 1991).[1] Though the theory is well supported by empirical tests, we make two improvements in the present research. First, we make the theory more precise: By taking into account actors' expectations, we generate ratio-scale predictions. Our refinement predicts exact exchange outcomes. Second, we make the theory more general: We test the theory in a new restricted information setting. Correct predictions for this setting mean we expand the theory's domain of applicability. The theory could then potentially subsume within its scope more social settings in the field — buying a house, for example. When negotiating for a house, a buyer may have little information about the profit her offer will provide the seller and little information about the number and nature of the seller's alternative offers.

Below we review network exchange theory and then describe a new integration of two previous lines of theorizing: the graph-theoretic power index (GPI) for network structures and the resistance model for bilateral negotiations. The extended theory also incorporates biases in actors' expectations induced by the number of their direct ties to other network positions. We first test this new version of the theory against five alternatives with data from full information experiments on four different exchange networks. Finally, the generality of the theory is tested by replicating one of the networks under a new restricted information setting.

Network Exchange Theory

Network exchange theory uses GPI to predict power and profit rankings in exchange networks. Its scope includes networks in which actors in directly connected positions engage in a series of negotiations over divisions of resource pools. Most interesting are networks in which prevailing structural conditions prevent some actors from exchanging at certain times. Such conditions foster power.

GPI: DETECTING STRONG POWER DIFFERENCES

Network exchange theory assumes that GPI can predict power and profit rankings in exchange networks by detecting a position's structural advantage or disadvantage. Here we present an intuitive explanation of how and why it works. The appendix to this article provides a more rigorous treatment.

GPI is calculated by counting paths out from a position in a network. It adds odd-length paths, which are advantageous, and subtracts even-length paths, which are disadvantageous. Odd lengths are advantageous because it means a position has alternatives or a partner's alternatives have alternatives and so on. For example, consider a simple network, three actors connected in a line: A-B-C. Actor A may exchange with B and B with C, but A may not exchange with C. If all actors may exchange only once in a round of bargaining, B gains power because the other actors compete for the single available exchange opportunity with B. We say that B has two 1-paths while A and C have only one. Even-length paths are disadvantageous because it means that a position's potential partners have alternatives that vie with A for exchange with the partners. Actor A has a 2-path through B to C. This is disadvantageous and subtracts from A's GPI score. It means that B has an alternative to exchange with A. Actor B has no 2-paths. Thus the GPI index for the positions in the three-actor line network are 0-2-0. B will have an overwhelming advantage in this network because B has a higher GPI score than A and C. We say that B has a strong power advantage.

Power changes dramatically with the addition of another actor to the above network, making a four-actor line, A-B-C-D (Willer & Patton 1987). Actor A now has a 1-path, a 2-path, and a 3-path. Its GPI score is 1. Actor B now has two 1-paths and one 2-path. Its GPI score is also 1. GPI predicts no strong power advantage for B in the four-actor line network. The reason for this is that the addition of a fourth actor gave A an additional, advantageous, odd-length path, and the addition also gave B an additional disadvantageous even-length path.

GPI extends this analysis to exchange networks of any size and density of relations by counting only nonintersecting paths leading away from a position. Only nonintersecting paths are counted because intersecting paths do not seem to change fundamental power relations in a network. For example, suppose we add a fourth actor, Z, to the three-actor line A-B-C. Actor Z is connected only to B. Actor A now has two disadvantageous 2-paths, one through B to C and one through B to Z. But because these 2-paths intersect at B, the additional disadvantageous 2-path makes no qualitative difference in A's relationship with B. B is still A's only possible exchange partner and B still has an alternative to exchanging with A. GPI ranks the two actors as in the three-actor line. B has a strong power advantage over A, but now GPI scores are more extreme. The GPI score for B is 3 because of the additional 1-path, while for A the GPI score remains 0.

GPI assumes that actors seek exchange with a potential partner who has a larger GPI only if no weaker alternative exists. (Here "to seek exchange" means to make competitive offers, a situation determined only by structural conditions.) Analysis of the Stem network (Figure 2) shows that both C_1 and C_2 have

GPI scores of 1. They will seek exchange with each other but not with B who has a GPI score of 2. The theory assumes that when this happens, GPI is recalculated among the resulting subnetworks of actors who mutually seek exchange. Two subnetworks in the Stem, A-B and C_1-C_2, result; all positions now have a GPI score of 1. No position is predicted to have an overwhelming power advantage in the Stem.

LIKELIHOOD OF EXCLUSION IN WEAK POWER NETWORKS

Markovsky et al. (1993) identify two kinds of power in exchange networks — strong and weak — distinguished by their structural bases and their consequences for exchange profits. The source of power in the two types of network is identical, however: exclusion from exchange. In strong power networks, one or more actors are excluded in every round of exchange by one or more others who, under given structural arrangements, need never be excluded. A position's GPI score encodes its potential to be excluded (or to exclude) relative to its partners. Immediate ties to partners — 1-paths — provide alternatives that enhance a position's potential to exclude or avoid exclusion. This holds for all paths of odd length. But partners' immediate ties to others — 2-paths — provide one's partners with alternatives to oneself and thus decrease a position's potential to avoid exclusion or to exclude. This holds for all paths of even length.

The idea behind weak power is that no position can consistently exclude another without incurring costs to itself (Markovsky et al. 1993). In most weak power networks, either all positions are prone to exclusion or no position is necessarily excluded.[2] That is, for each position there is some outcome in which it is excluded from exchange or it is possible for all positions to be included in exchanges simultaneously. GPI registers these conditions by assigning the same score to all positions and thus predicts no strong power differences. However, GPI measures a position's susceptibility to exclusion on the basis of the pattern of ties alone. Markovsky and associates' (1993) weak power extension to network exchange theory takes account of other factors, in particular the pattern of activity in these ties that could induce differential susceptibility to exclusion.

In strong power networks, profit distributions approach maximum differentiation where the advantaged actors earns between 90% and 100% of available profit. In contrast, profits in weak power networks are more sensitive to actors' strategies, and profits from exchange stabilize well short of maximum differentiation. Generally, the advantaged actor in a weak power network earns between 51% and 75% of available profit. The different levels of profit differentiation between strong and weak power networks reflect the different bases for excludability — i.e., the pattern of ties versus the pattern of activity in those ties.

For example, the Stem is a weak power network. With strong power, profit distributions approach maximum differentiation. That is, if actors negotiate over the allocation of a resource pool containing P units, profits for high-power actors will approach P and those for low-power actors will approach zero. In contrast, profits in weak power structures are more sensitive to actors' strategies, and profits from exchange stabilize well short of maximum differentiation,

e.g., at $(P/2) + 1$ for the advantaged actor, and $(P/2) - 1$ for the disadvantaged actor. In general, structurally disadvantaged actors face more exclusions from exchange than advantaged actors, and when excluded they respond by making offers that slightly favor actors in advantaged positions.

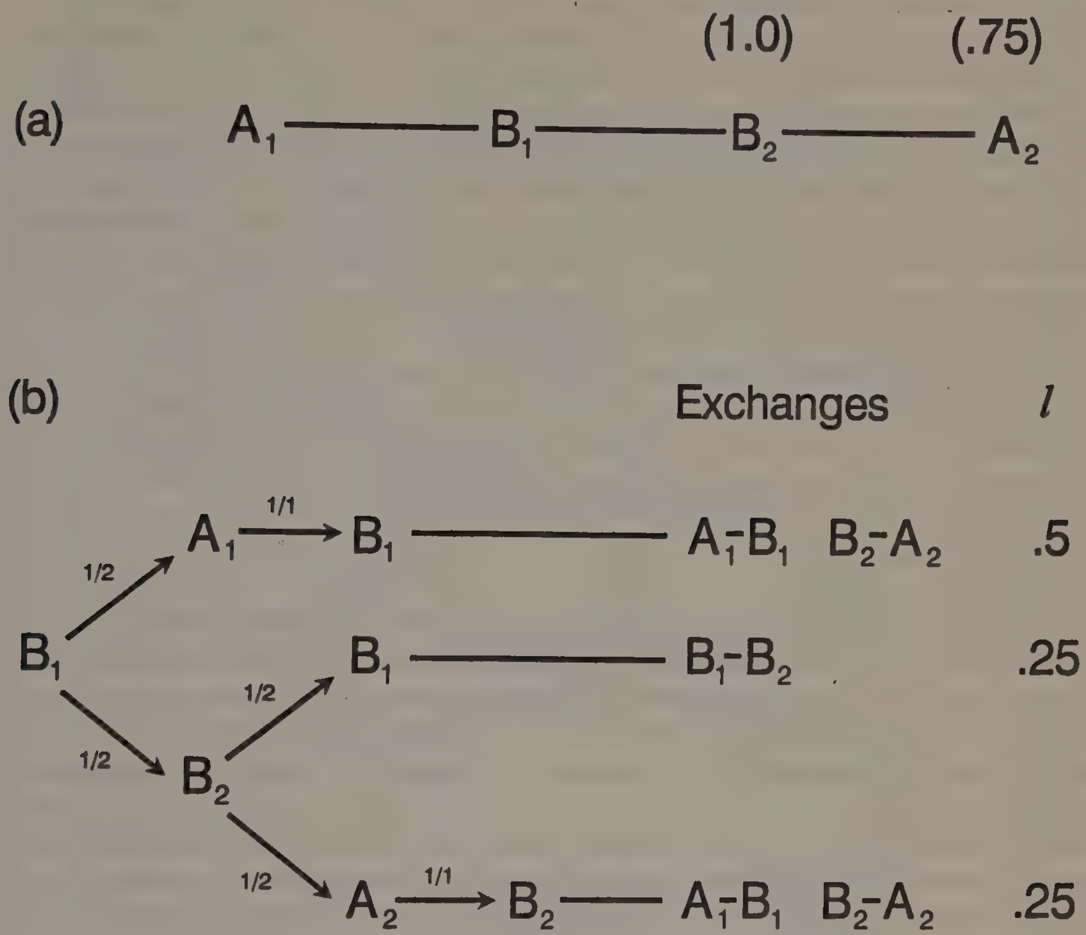
The Markovsky, Willer & Patton (1988) GPI model correctly identified strong power structures in all its tests. That is, (1) unequal GPI values correctly anticipated unequal profits, (2) such profit inequalities were relatively large, and (3) whenever profit levels were equal, then so were GPI values. However, Skvoretz and Willer (1991) found that actors' profits may differ even when their GPIs are equal. Prediction of these profit differences required a second step taken in Markovsky et al. (1993): When all GPIs are equal, each position's likelihood of being included in exchanges, l (or of being excluded, $1 - l$), is calculated under the assumption that actors have no preferences among partners.[3] Then, in an i-j relation, actor i has weak power over j only if GPI values are equal and if $l_i > l_j$ (see Appendix). Otherwise, $l_i = l_j$ and i and j are equal in power. Thus, likelihood of being included detects weak power in networks where GPI detects no strong power differences. The analysis of l_i is not applicable in strong power structures, but rather GPI is used to predict the very robust profit differences that occur there. In summary, we detect power differences in two steps. First, GPI is applied to find strong power differences. Second, where no strong power differences are found, likelihoods of being included are calculated to assess any weak power differences that may exist.

As an example, consider the four-actor line of Figure 1A. GPI = 1 for all positions. Because positions have equal GPI, no strong power differences exist. Therefore, we turn to calculation of l_i to check for any weak power differences. In this network, A_1 can negotiate and exchange with B_1; B_1 may do so with A_1 or B_2; and so on down the line. Assuming that each actor can exchange only once per round and is indifferent to with whom, either B has a .5 probability of seeking exchange with an A and a .5 probability of seeking exchange with the other B. The probability that an A will seek exchange with a B is 1.0 because the A's have no alternatives.

Figure 1b shows a probability tree used to calculate l for each position. Each branch of the tree shows an exchange-seek and its associated probability. The Exchanges column shows mutual exchange-seeks and the l column shows the product of branch probabilities leading to each possible exchange. The likelihood of exchange between two actors is the sum of the probabilities associated with their mutual exchange-seeks in the Exchanges column. An actor's likelihood of being included in any exchange is the sum of the probabilities associated with all exchanges that include that actor. From the tree we derive that $l_{AB} = .75$ for both A-B pairs; $l_{B_1B_2} = .25$; $l_A = .75$; and $l_B = 1$.

The foregoing procedure generalizes to more complex networks and to networks where actors can exchange more than once in a round of bargaining (Markovsky 1992; Markovsky et al. 1993). The distinction between single and multiple exchange is theoretically important because it determines the pattern of exclusions. This was first shown by Markovsky, Willer & Patton (1988), who demonstrated that changing the number of exchanges allocated to positions altered the power exercised in every relation in the network. More recently, Skvoretz and Willer (1993) showed similar consequences for a new array of

FIGURE 1: Four-Actor Line and Its Likelihood



networks. A real-world example will show why altering the number of exchanges has such a great impact on power. Auctions are a network exchange process. The seller has only a finite amount of time to sell any and all items at the best possible price. The finite period for the auction corresponds to a round of negotiation, that is, the period in which exchanges involving existing pools of resources must be completed. A resource pool might consist of a single item or several identical items, or there might be several resource pools containing a wide variety of items as in a real auction. Different items may appeal to some or all of a wide variety of bidders, which determines the pattern of relations in the exchange network.

Consider two simple auctions. In both we have one seller and two buyers. In the first auction, there is only one item, an antique lamp, of interest to both buyers. In the second auction, there are two such lamps. In the first auction, because there is only one lamp, the seller and each of the buyers can make only one exchange during the auction but one of the buyers will be excluded from exchange. As the bidding opens, each buyer must try to provide the seller with the more lucrative offer. Because each buyer wants to make the higher bid, however, the offers "ratchet" higher and higher and become increasingly

favorable to the seller. Eventually, one buyer backs down and the seller accepts a tidy profit.

In the second auction, each buyer (who wants only one lamp) still seeks exchange with the seller. But because there are two lamps, the seller now may exchange twice during the auction period — once with each buyer. The buyers do not need to outbid each other in order to obtain the lamp they desire, and so there is no structurally induced exclusion from exchange and no price ratcheting. Ironically, the combined prices for the two lamps may end up being less than what the seller could have obtained by placing only one up for auction. Varying the number of permitted exchanges emphasizes the importance of exclusion as the generator of structural power in networks, as well as the generality of the theory (Skvoretz & Willer 1991).

Despite the theory's generality, the structurally induced exchange likelihoods that it generates have been used to predict only the ordering of exchange outcomes across positions in weak power networks. We now turn to the problem of predicting exact exchange rates in weak power networks.

Profit Expectations

Network exchange theory depicts the generation of profit differentials in networks as an almost purely structural phenomenon. Profit differentials arise from differences in avoiding exclusion and from the sheer pattern of ties. The cognitions of actors play very little role in explanation. The strategy has been fruitful for two reasons. First, structural factors are often sufficient to predict accurately simple ordinal differences in earnings. Second, strong power networks have played an important developmental role in network exchange theory. In strong power networks, structural determinants are so powerful — as indicated by consistently extreme profit differentials — that actor cognitions can introduce only minor variation at best. But the study of weak power networks demands that more sources of variation be taken explicitly into account. In particular, we hold that more precise prediction requires that we extend network exchange theory to incorporate actors' profit expectations. We concentrate on possible sources of actors' expectations that might develop from initial network conditions and ongoing feedback that might result from them because our goal is to predict actor behaviors and exchange outcomes on the basis of initial conditions.

In exchange networks, initial conditions and ongoing feedback provide actors with information that they can use to estimate their potential profits, e.g., the maximum amounts they could hope for, the minimum outcomes they fear, and the profits that might reasonably be expected to obtain. To the extent that such profit expectations are affected by situational factors, negotiations will be modified and, in turn, exchange profits will be affected. The sensitivity of the negotiation process to contextual information should be especially evident in experimental settings. There, the simplicity of exchange conditions focuses attention on whatever minimal information is provided.

Our strategy is thus to employ a formal model that (1) accounts for the effects of profit expectations, (2) can be readily extended to accommodate situational factors that modify such expectations, and (3) may be integrated

seamlessly into existing network exchange theory. In line with the existing theory's emphasis on the importance of exclusion and excludability, we assume that a situational factor also modifies actors' profit expectations. It would be surprising if it did not. We have all been excluded from social exchange at some time in our lives whether it was not being invited to a party or having our application to a university rejected. These events certainly affect our cognitions. The following section incorporates the actors' experience of patterns of exclusion into our model. The second factor we will highlight in the section entitled Expectations and the Number of Direct Ties. Previous research (Marsden 1983) suggests the significance of the number of direct relations an actor has in the network. But more important, the number of direct relations is a highly salient and immediately apprehensible feature of the actor's environment even under conditions of highly restricted information.

EXPECTATIONS AND RESISTANCE

The resistance model (Heckathorn 1983; Willer 1981) predicts rates of exchange based on each actor's beliefs or expectations about his or her own best and worst possible outcomes. Specifically, it assumes that the point at which actors agree to exchange is determined by balancing two interests: (1) Actors aspire to obtain the greatest possible profit from exchange. This is their "best hope." (2) Actors seek to avoid the worst possible outcome. This is their "worst fear." To decide whether to agree to exchange, actors balance their desire for maximum profit against their fear of receiving no profit or the profit that results if no exchange occurs.

In dyadic exchange, the scope of the resistance model overlaps the Nash equilibrium (1950, 1953), and under certain conditions the predictions of the two are the same. However, the scope of the resistance model is broader, extending beyond the dyad. It has been successfully applied to a wide range of exchange networks (Brennan 1981; Willer 1987; Willer, Markovsky & Patton 1989). Its predictions have also been shown to hold in cross-national experimental research (Willer & Szmatka 1993).

We use the resistance model to formalize our claims about the effect of profit expectations on negotiations. It specifies how negotiators arrive at agreements through each actor's beliefs or expectations about his or her own best and worst possible outcomes. Implicitly, the model may be interpreted as generating profit expectations for negotiating actors at a particular point in time, and these expectations then determine whether actors accept or reject offers.

Let P_i represent i's profit from exchange, M_i is i's maximum expectation or best hope for profit from exchange, and C_i is i's worst fear or "conflict outcome."[4] Actor i's interest in gaining his or her maximum expectation is $M_i - P_i$ and in avoiding his or her worst fear is $P_i - C_i$. His or her resistance to a given exchange profit, R_i, is the ratio

$$R_i = \frac{M_i - P_i}{P_i - C_i}$$

Resistance is the ratio of $M_i - P_i$, the actor's interest in gaining a better outcome, to $P_i - C_i$, the actor's interest in avoiding disagreement. Because the ratio is

small for favorable settlements and large for unfavorable settlements, resistances of two actors in an exchange relation vary inversely for a given settlement. Network exchange theory asserts that agreements to exchange occur when actors' resistances are equal. Thus, compromise occurs when:

$$\frac{M_i - P_i}{P_i - C_i} = \frac{M_j - P_j}{P_j - C_j}$$

This is the *equiresistance* solution.[5] Knowing the number of resource units in the pool (P) such that $P = P_i + P_j$, we may algebraically solve for the values of P_i and P_j. More than a decision strategy, resistance is conceived as a potential limit of power use when actors use the best available strategies. Consequently, it holds promise of more general applicability than any particular decision strategy.[6]

To predict negotiation outcomes, M_i and C_i must be determined for each actor. In strong power structures such as A-B-A, we assume that each A's maximum expectation (M_A) is initially at or near P. However, given that B seeks only one exchange per round, M_A declines as A's are consistently excluded. M_A may begin at or near P, but this best hope will approach zero as exchanges continue to yield ever-declining profits. In contrast, because B has no rivals, M_B remains close to P. Over a series of rounds all profit gravitates to the central, B, actor (Willer & Markovsky 1993).

C_i is determined by a position's best alternative (Willer 1987). Strong power structures such as A-B-A are characterized by bidding wars between rival A's. Thus, when B negotiates with one A, C_B is the last offer from the other A. As the two A's bid, however, C_B increases toward P. Because A has no alternative, C_A stays at zero. The result over a series of exchange rounds (derivable from the equiresistance model) is that P_A approaches zero and P_B approaches P (Willer & Markovsky 1993).

In weak power structures, an equilibrium exchange rate is reached such that neither actor gains the maximum available profit or is forced to accept almost none. To use resistance to predict exchange rates in these structures, it is necessary to determine the value of actors' maximum expectations and conflict outcomes, M_i and C_i, at equilibrium.

In weak power structures, actors do not initially have a realistic basis for estimating their maximum expectations and conflict outcomes, M_i and C_i. As a result, their initial expectations may be either optimistic or pessimistic. However, all equiresistance solutions assume that actors have expectations and that over a series of exchanges they come to be more or less realistic. Thus, we assume that during the interaction process actors learn more realistic expectations. For example, when initial expectations are too optimistic, actors are excluded by others and eventually adjust their expectations downward. If expectations are initially too pessimistic, then actors always gain agreements and eventually adjust their expectations upward. As a result, expectations become increasingly realistic. This model of actors' responses to inclusion and exclusion conforms to the scope specifications for offer adjustments first described by Markovsky, Willer & Patton (1988). Here, however, we make the more specific suggestion that actors adjust their offers in response to changes in their profit expectations.

Below, we offer a two-part solution for the problem of predicting exchange rates in weak power networks. First we predict the equilibrium exchange rate. If actors adjust expectations as we assume they do, then expectations will come to correspond more closely to actors' likelihoods of being included in exchange. The result will be a fairly stable equilibrium exchange rate.

Second, we note that initial rates may not be like equilibrium rates. Because actors do not initially know their likelihood of being included in exchange, their expectations for maximum profit and their worst fears must have other bases. Initial expectations may or may not be realistic. We also offer a simple model for initial expectations from which actors move, as a consequence of interaction, toward equilibrium. These initial expectations likely have an enduring impact on exchange rates. In a final step, we complete our theory by combining the model of initial expectations with equilibrium exchange rate predictions. In effect, our predictions assume that actors' beliefs remain biased to some degree by initial expectations.

RESISTANCE AND THE LIKELIHOOD OF BEING INCLUDED

We approach the problem of specifying the value of conflict outcomes and maximum expectations in weak power networks by first identifying theoretical restrictions for C_i and M_i. Within these theoretical limits, we then assume that C_i and M_i are proportional to an actor's likelihood of being included in exchange. That is, an actor's expectations for profit, her or his worst fears and best hopes, depend on how often she or he expects to be included in profitable exchange. The assumption of simple proportionality between inclusion expectations and likelihood of being included results in a modified resistance equation that can be used to make exact predictions of exchange rates between actors in weak power networks.

The conflict expectation, C_i, depends on the actor's expectations regarding available exchange alternatives, as noted above. For example in the four-actor line network, A_1-B_1-B_2-A_2, if B_1 knows that A_1 will agree to an equal division of profit at 12:12, then B_1 will not accept 11 from B_2. However, as Yamagishi and Cook (1990) noted, actors are not always certain of their alternatives. Nonetheless, weak power limits the range of conflict outcomes: the lower limit is zero — the amount an excluded actor receives — and the upper limit is half the total resource pool, as we will next explain.

In all weak power networks, actors cannot consistently exclude others from exchange without themselves suffering losses.[7] Still assuming 24-unit resource pools, this can be illustrated using the four-actor line. Suppose that C_{B1} is greater than 12, an equal division of the profit pool. This means that B_1 would refuse offers of less than 13. That is, B_1 will refuse an equal division of profit and hold out for more. Because B_1 cannot consistently exclude B_2, however, B_2 will not be penalized for refusing to exchange with B_1 at this rate. As long as B_1 demands 13 from B_2, B_2 effects a temporary break in the network by exchanging with A_2 at no worse than 12:12. For B_1 to ever reestablish the possibility of exchanging with B_2 (and thus reestablish weak power over A_1), C_{B1} must be reduced to 12, an equal division of the profit pool, or lower. Therefore, it is generally true that the "conflict" or worst-fear outcome in weak power

structures is limited to the range from zero to half the size of the resource pool. Similarly, M_i is restricted to the range 12-24. M_i may be close to the total resource pool when exchange begins, but it cannot go below half the pool. Half the pool is always a competitive offer because no actor is ever consistently excluded.

At issue now is how to determine C_i and M_i from initial conditions within these theoretically determined ranges. We do so using l, the likelihood of being included, as derived from weak power calculations (Markovsky et al. 1993). Two assumptions integrate resistance with Network Exchange Theory. First,

Equiresistance Assumption: In weak power relations, actors' profits approach equiresistance solutions over a series of exchanges.

To the extent that actors use effective strategies to seek maximum profits, their profits should conform to the resistance predictions. For example, the behavior of more experienced actors or those with better training should conform more closely to resistance predictions than the behavior of less experienced or less well trained subjects.[8] The idea that resistance predicts a profit limit reached at equilibrium after a series of exchanges allows application of the theory to exchange situations in which actors have different amounts of information and training.

Second, we theoretically link inclusion probabilities to actors' conflict outcomes and best-hope outcomes in the resistance equation. We assume that an actor's perceived best-hope and conflict outcomes are proportional to that actor's likelihood of being included in exchange. For example, an actor who is consistently included and who makes a profit would resist offers that are lower than she or he is accustomed to receive. The fact that she or he is very likely to be included in exchange has increased her or his point of conflict. Conversely, a frequently excluded actor, accustomed to receiving no profit much of the time, would accept a low offer. The actor's low likelihood of exchange has reduced her or his point of conflict. The same argument can be made for actors' best hopes. Actors consistently included in exchange should have higher aspirations for profit than should actors consistently excluded from exchange.

Markovsky et al. (1993) demonstrated that likelihood of being included (l_i) ranks the power of positions in weak power networks. On this evidence, we assume that l_i will successfully rank power positions even where power differences are very small. Further, we assume that larger differences in l_i identify larger differences in power between positions. That is, we assume that weak power is proportional to a position's l_i. If as we suggest, this occurs because likelihood of being included acts on points of conflict (C_i) and best hopes (M_i), then setting C_i and M_i proportional to l_i should provide a good indicator of an actor's power.

The following resistance-likelihood assumption expresses the idea that C_i and M_i are proportional to l_i within their respective ranges. Our analysis demonstrated that C_i is limited to a range between zero and half the resource pool. Similarly, M_i is limited to a range from half the resource pool to the entire pool. This assumption predicts that the difference in profits for high-power and low-power actors in a weak power relation depends upon their likelihoods of being included in exchange.

Resistance-Likelihood Assumption: The higher an actor's likelihood of being included in exchange, (1) the higher the actor's perceived conflict outcome, C_i; and (2) the higher the actor's maximum profit expectation, M_i. Formally,

$$C_i = \frac{P}{2} l_i \qquad (1)$$

$$M_i = \frac{P}{2}(l_i + 1) \qquad (2)$$

In words, the perceived worst-case exchange outcome (C_i) is a fraction of half the pool, and that fraction is larger for higher likelihoods of being included (l_i), and smaller for lower l_i. As explained above, C_i is limited to at most half the resource pool in weak power situations. Equation (1) expresses the assumption that C_i is proportional to l_i and ranges between zero and half the resource pool.[9] Similarly, an actor's maximum expectation for profit (M_i) is half the pool plus a fraction of half the pool. We also showed that M_i is restricted to be at least half the resource pool but not more than the total pool, P, in weak power situations. Equation (2) expresses the assumption that M_i is proportional to l_i on that range. In weak power situations at equilibrium, it follows from equations 1 and 2 that M_i is a direct function of C_i, i.e., $M_i = C_i + P/2$. This feature pays considerable dividends in simplifying calculations and serves as a plausible assumption about perceived best- and worst-case outcomes.

The resistance model, resistance-likelihood assumption, and a little algebra yield the following prediction:[10]

$$P_i = \frac{(P + C_i - C_j)}{2}$$

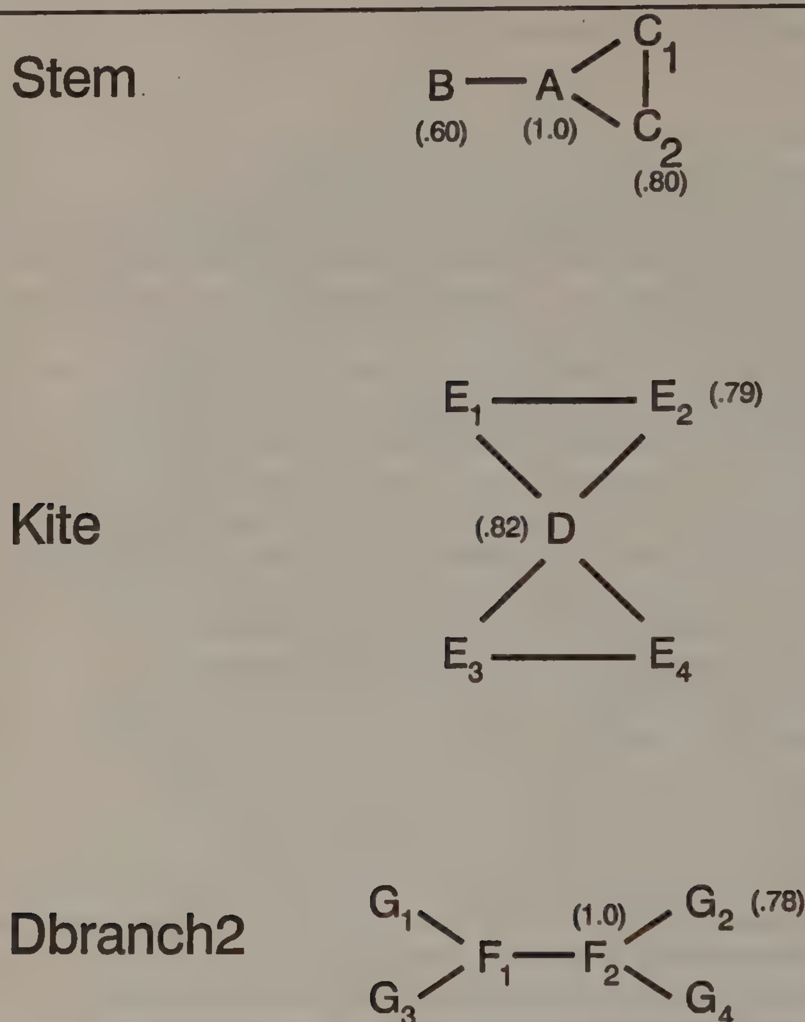
$$P_j = P - P_i.$$

In the Stem network (Figure 2), for example, likelihoods of being included for positions A and B are $l_A = 1$, $l_B = .6$. Assuming a typical 24-point resource pool, P, we can calculate profit distributions for the A-B relation. First, by the resistance-likelihood assumption, $C_A = (24/2) \bullet 1 = 12$, and $C_B = (24/2) \bullet .6 = 7.2$. Next, profits are calculated to be $P_A = (P + C_A - C_B)/2 = 14.4$, and $P_B = P - P_A = 9.6$. Profits for any position in any weak power network can be predicted by this method if the network structure and total value of each profit pool are known. Likelihood of being included indicates the structural power advantage of a network position. By balancing two competing motives — the desire to increase profit and the need to reach agreement and avoid exclusion — actors reach an equilibrium level of profit that is proportional to their relative likelihood of being included.

EXPECTATIONS AND THE NUMBER OF DIRECT TIES

In addition to the factors that we have previously related to profit expectations, there is good reason to assume that the number of an actor's direct relations in the network may also play an important role. In the parlance of network analysis, this refers to the actor's degree. Marsden (1983) employed a similar idea in his theory of power in exchange networks. Where t_i is the number of actor *i*'s direct network ties, he defined $\log(t_i/t_j)$ as one of several factors affecting *i*'s "price-making" behavior in exchanges with *j*. Although some of his

FIGURE 2: Three Weak Power Networks and Likelihoods of Inclusion



model's predictions diverge from experimental test results, the notion that degree biases actors' price negotiations in weak power situations has not been tested directly and may still be sound. Further justification is found in the judgment heuristics literature (e.g., Kahneman, Slovic & Tversky 1982). In a wide variety of judgment contexts, informational anchors have been shown to bias judgments of such properties as magnitude, numerosity, value, weight, color, loudness, pitch, etc. For example, a contrast anchor effect occurs when yesterday's 95 degree temperature (the anchor) makes today's 78 degrees seem cool. Today's judged temperature is biased away from yesterday's. The assimilation anchor effect is often found in negotiation settings where an initial offer (the anchor) may be blatantly unrealistic, but subsequent offers and counteroffers are still pulled toward that initial offer. (For more detailed examples and applications see Helson & Kozaki 1968 and Kahneman, Slovic & Tversky 1982.)

Assuming that human actors cannot fully evaluate the ramifications of their location in a network structure — especially when lacking systemwide information — it is reasonable to presume that information of a more localized

nature becomes especially salient. The number of one's direct ties is just such a piece of information. An actor with more direct network relations will probably expect to have more successful negotiations than an actor with fewer direct relations. It seems plausible that an actor with many potential exchange partners would expect a better chance of being included in exchange than would an actor with only one or very few exchange partners. Of course, network exchange theory shows that this expectation is not necessarily justified. The extent to which having alternative partners can benefit an actor depends on broader network patterns, e.g., the alternative relations of each of one's alternative relations. Nevertheless, an actor's degree is a highly salient piece of information in network exchange contexts and should thus bias profit expectations via an assimilation effect.[11] This idea is captured in the following:

Degree Assumption: The higher an actor's degree, the higher the actor's expected profit.

An actor with higher profit expectations is assumed to adopt a tougher bargaining stance, e.g., to make lower offers to others, and to have higher thresholds of acceptability for incoming offers. If negotiating actors have equally high degrees, however, degree would not provide special advantages to either. Therefore, our index of relative degree (d_{ij}) for actor i in the i-j relation must actually be based on the relative number of ties (t) for each actor. This is accomplished in the formula

$$d_{ij} = \frac{t_i}{(t_i + t_j)}.$$

This specification standardizes the index to a 0-1 scale, a useful property when we combine d with other components of the theory. As a structural measure, d_{ij} does not depend on actors' knowing one another's degree. It captures what might be called the expectation advantage of one actor relative to another — a condition that will then manifest itself in the dynamics and outcomes of the negotiation process.

RESISTANCE AND DEGREE

Degree is assumed to bias profit expectations. Therefore, in the model, we incorporate the relative degree index as a biasing factor for best-hope and worst-fear outcomes. Given that we have already defined M (the maximum hoped-for profit) to be a function of C (the conflict or worst-fear outcome) in weak power networks, we need to show only how relative degree affects the latter. We assume that, in the same manner as the likelihood of being included, the higher the degree, the greater the inflation of the actor's worst-fear outcome:

Resistance-Degree Assumption: The higher an actor's relative degree, the higher the actor's perceived conflict outcome.

Combining this assumption with the resistance-likelihood assumption yields

$$C_{ij} = \frac{P}{2} l_i d_{ij}$$

We now subscript C by both i and j, because the inclusion of the biasing factor, d_{ij}, in the equation implies that i's expected conflict outcome may be different for each actor with whom i negotiates.[12] Again using the Stem network as an

example, and using the earlier formulas for P_i and P_j with a 24-point resource pool, $d_{AB} = 3 / (3+1) = 3/4$. Substituting the values for variables in the equation, $C_{AB} = 24 / 2 \cdot 1 \cdot 3 / 4 = 9$; $C_{BA} = 24 / 2 \cdot .6 \cdot 1/4 = 1.8$. Then we can solve for the prediction that incorporates both resistance and relative degree into the GPI model, $P_A = (24 + 9 - 1.8) / 2 = 15.6$, and $P_B = 8.4$. We label this elaborated model GPI-RD.

Higher relative degree is thus assumed to bias the effects of the likelihood of being included. However, because advantages in relative degree are based only on actors' expectations and not on actual structural advantages, there is a potential cost to actors who try to exploit degree advantages. Though an actor with more potential exchange partners may be a tougher negotiator and receive more resources from completed exchanges, such exchanges may be less frequent than for actors of lower degree. This is because higher-degree actors are prone to tougher negotiation than is warranted by their actual structural positions. The result would be higher profit when exchange occurs, but a higher likelihood of being excluded from exchanges. In addition, actors with fewer potential exchange partners still seek more favorable alternative agreements whenever possible. In networks where all actors can be excluded — l_i less than 1.0 for all actors — a situation could arise in which an advantaged actor is excluded from exchange so frequently that she or he can actually acquire fewer resources than the disadvantaged actor over a series of exchanges. Markovsky et al. (1993) report this phenomenon in the "Kite" network (Figure 2). Subjects in the D position have an advantage over those in E positions both in likelihood of being included, .82 to .79, and relative degree, 4/6 to 2/6. For experienced subjects, despite the very small advantage in likelihood of being included, D position subjects averaged a 14:10 profit point advantage in exchanges with E partners. However for D these victories were Pyrrhic. Ds were excluded — and earned no profit — on 41% of rounds, while Es were excluded on only 15% of rounds.[13] Es preferred to exchange with each other rather than the more aggressive D. It seems strange that a high-power actor can actually earn less overall than a low-power partner. Incorporating degree into the theory explains this result — a result that previously had been considered anomalous. Though frequency of exchange can also be predicted using likelihood of being included and degree (Skvoretz & Lovaglia n.d.), here we focus on testing the model's predictions for exchange profits.

Method

We tested the model's predictions using experimental data from four networks. Subjects in the test setting had full information on the network structure and on all other actors' exchange outcomes. We also replicated the test on one of these networks using a restricted information setting where each subject knew only his or her own dealings with potential exchange partners and his or her own profits from exchanges.

EXPERIMENT 1: "FULL INFORMATION" NETWORKS

Skvoretz and Willer (1991) described in detail the experimental setting used for these tests. Custom software ("ExNet") was used to configure networked PCs into "virtual" exchange networks. ExNet can establish networks of any shape and size, limited only by the number of PCs. Subjects in experiments know the network structures and their positions within them. In an initial session, assistants explain how to make offers and counteroffers, what it means to divide the 24-point resource pool, the dollar value of profit units, and that exclusion from exchange yields no points for that round. A practice session familiarizes subjects with the operation of the system. Subjects then return at a later date to participate in the actual research. In the experiments, each subject rotates through all network positions, negotiating for a total of four rounds at each position. The computer records the timing and content of all communications.

We investigated four experimental structures of theoretical import (Figure 1a and Figure 2). The Double Branch 2 is a simple weak power network that allows two positions multiple exchanges. It was converted from a strong to a weak power structure via a two-exchange rule: The central Fs could exchange up to twice per round. Actors are only allowed one exchange per round in the other structures. These networks also have interesting theoretical properties that warranted our attention: The four-line is the simplest weak power network; the Stem and Kite sparked a debate between competing research programs over the validity of the GPI (Markovsky, Willer & Patton 1990; Yamagishi & Cook 1990); and in the Kite, no position is guaranteed inclusion.

Five alternative theories are capable of generating ratio-scale profit predictions for positions in weak power networks. Most of these were presented in a special issue of *Social Networks* (Bienenstock & Bonacich 1992; Cook & Yamagishi 1992; Friedkin 1992; Skvoretz & Fararo 1992). Bienenstock and Bonacich's core theory takes a game-theoretic approach. Cook and Yamagishi's equidependence theory extends the power-dependence program originated by Emerson (1962). Friedkin's expected value theory developed out of his work on network analysis in general. Skvoretz and Fararo (1992) apply Coleman's (1973, 1990) rational exchange theory to these exchange structures. In addition, Skvoretz and Willer (1993) present exchange resistance theory, a model that incorporates resistance into the GPI but does not include degree. These five theories are briefly described below. Except for Coleman's rational exchange model, Skvoretz and Willer (1993) provide details for using each to calculate predicted exchange outcomes.

CORE THEORY

Core theory (Bienenstock & Bonacich 1992, 1993) models network exchange in terms of cooperative game theory.[14] Three rationality criteria establish the "core" of an exchange network. First, agreements are individually rational when each actor's profit is equal to or greater than the profit that could be earned by not exchanging. That is, actors are assumed to exchange only when they receive at least as much profit from agreement as they do from being excluded from exchange. Second, agreements are rational for the coalition of exchange partners

when the sum of the profit of both actors is at least as much as could be obtained if they exchanged with other partners. Third, group (or network) rationality obtains when the total profit of all positions of the network is at least as large as the total profit available from some other pattern of exchange agreements.

The core of an exchange network usually narrows the range of preferable exchange rates but does not necessarily predict a single ratio-scale exchange outcome. To make predictions when this occurs, we follow Skvoretz and Fararo (1992) in assuming that each core outcome is equally likely and then average the payoffs to various positions. In some cases, networks contain no core at all, which precludes a prediction.

RATIONAL EXCHANGE THEORY

Coleman's (1973, 1990) rational exchange model is not easily applied to our exchange networks because it operates under scope conditions that differ from those of other network exchange theories. It assumes, for instance, that every actor may exchange with any other actor in a network. Marsden's (1983) model solves the problem by adding network restrictions and a variety of additional assumptions. However, he noted that some predictions ran contrary to data previously published. Coleman partially solved the problem by assuming that there are transaction costs between actors. When transaction costs are high, they effectively prohibit exchange from occurring, thus setting the stage for structural power to emerge. As it stands, however, Coleman's conception of power is not relational. Power manifests in an actor's resource holdings rather than in her or his relative ability to extract resources through exchanges. Also, it does not address the situation where one actor's exchange is contingent on whether another actor exchanges. Skvoretz and Fararo (1992) have modified and added assumptions to Coleman's theory to make it applicable in the kind of exchange networks discussed here. Predictions for the Coleman model are based on their analysis. A technical description of the method and a computer program that calculates predictions are available from the second-named author on request.

EQUIDEPENDENCE THEORY

Equidependence theory (Cook & Yamagishi 1992) assumes that actors compare how much they will receive in exchange with a potential partner against how much they could get in exchange with some other partner. The difference between what an actor can obtain from this exchange relation and that of an alternative relation is deemed to be the dependence of the actor on the potential partner. This comparison process goes on simultaneously with all an actor's direct ties. Actors are assumed to exchange at a point where their dependence on the relation is equal to the dependence of their potential partner. In other words, exchange occurs when actors are equidependent, and each actor can get no more profit by some alternative exchange. An actor's reward from exchange is given by the equation $R_{ij} = (P_{ij} + A_{ij} - A_{ji})/2$, where R_{ij} is the profit that actor i obtains in exchange with partner j; P_{ij} is the pool size, and A_{ij} is the best alternative available to i.

Cook and Yamagishi use the example of two actors, *i* and *j*, negotiating over a 24 point resource pool. Actor *i* has another partner who offers *i* 10 points, while *j* has no other partner. If *i* and *j* were to divide the pool at 13 for *i* and 11 for *j*, actor *i* gets 3 points more than his or her best alternative (10), but *j* gets 11 points more than his or her alternative (zero). Thus equidependence between *i* and *j* has not been reached. Negotiation is assumed to continue until *i* gets 17 points and *j* gets 7 points. Here, equidependence has been attained because both *i* and *j* get 7 more points from exchange with each other than they would outside the *i*-*j* relation. Actor *i*'s power is defined as the maximum profit she or he can obtain from any partner.

EXPECTED VALUE THEORY

Friedkin's (1992) expected value theory first identifies all subnetworks that can result from various patterns of exchanges. Expected values are predictions about a structure's outcomes determined by weighting the value of a predicted outcome by the probability of its occurrence. For example, in the four-actor line network, A_1-B_1-B_2-A_2, there are two possible exchange patterns. Either each B exchanges with its related A, or the two Bs exchange with each other. To make predictions from initial conditions, Friedkin assumes that either pattern is equally likely to occur. Actor *i* is dependent on actor *j* if failure to exchange with actor *j* results in *i*'s exclusion from exchange. Dependency is an actor's likelihood of being excluded, calculated over all possible exchange outcomes. An offer-making function translates a particular degree of dependency into an offer to a potential exchange partner. Predicted earnings from exchange are a function of reciprocal offers modified by compromises when offers are inconsistent.

EXCHANGE RESISTANCE THEORY

Skvoretz and Willer (1993) use the likelihood of being included and the resistance model to make "baseline" predictions for exchange in both strong and weak power structures. Their goal is a simple formula that can be used in a single step to yield predictions in all exchange networks. We give their formula for predicting the profit of actor *i* using the notation for our own resistance-likelihood assumption:

$$\frac{\ln(M_i - P_i)}{\ln(P_i)} = \frac{l_j}{l_i}$$

They assume that conflict points and maximum expectations for profit for all actors remain constant. The conflict point for all actors is assumed to be zero, and their maximum expectation for profit is assumed to be the entire profit pool. A power function is then applied in which the difference between the maximum expectation for profit and what an actor would receive from an offer is raised to the power of that actor's likelihood of being included in exchange. The equiresistance equation can then be reduced to the above equation using natural logarithms.

All five theories make ratio-scale predictions for at least some of the network relations we examined. Below we compare our predictions to those of the five alternative theories and to experimental results.

Results

Four groups of subjects participated in Stem and Kite experiments, and five groups in the four-line and Double Branch 2. All were university students who signed up to participate in paid experiments. In Table 1, the column headed GPI-RD shows the predictions for our new model that integrates GPI, resistance, and degree. The column headed GPI-R shows the predictions for our resistance model without the biasing effects of degree. Also shown are the predictions from five alternative models and the observed means by network relation. Because we assume that profits approach predicted equilibria over a series of negotiations, we use data from the final experimental periods.[15]

Observed profits conform well to GPI-RD predictions. The largest discrepancy between a predicted and an observed value is less than one profit unit. One-sample t tests determined the probability that the differences between predictions and observed means were due to chance. No GPI-RD prediction differed significantly from its corresponding observed mean at or below the .40 probability level. The smallest probability that a prediction did not differ from the observed value was .47 for the Kite network. Two predictions from Coleman's (1973, 1990) rational exchange model, while not as close, were better than those of other alternatives. The probability that rational exchange predictions did not differ from observed values was .13 for the four-line and .11 for the Stem. In contrast, significant differences were found between observed means and the predictions of other alternative theories. For the Kite network, Skvoretz and Willer's (1993) exchange resistance model, GPI-RD, Cook and Yamagishi's (1992) equidependence model, and Friedkin's (1992) expected value model all make acceptably close predictions. However, both the equidependence and expected value models predict no difference in power between D and E actors in the Kite — a difference that did occur in an empirical test and was statistically significant (Markovsky et al. 1993).

Establishing a GPI-RD prediction for the Double Branch 2 network requires some interpretation. Calculating degree as for a one-exchange network, Fs have three direct ties and Gs have one. The GPI-RD model then yields predicted profits for F of 15.33, about 1 profit point away from the observed value. A one-sample t test also finds this difference significant: $t(63) = 2.29$, $p = .03$. However, degree can be calculated differently when two exchanges are allowed. Fs exchange twice per round while Gs exchange only once. At the beginning of a round, F has three potential exchange partners. If F's first exchange in a round is with a G, then there are two actors left with whom to attempt a second exchange. If Fs first exchange with each other, then they have three actors left with whom to attempt a second exchange. Fs then have either five or six direct ties while Gs have only one. If F's first exchange is with a G on two-thirds of the rounds (i.e., F is indifferent between the other F and its two Gs), then Fs effectively have 5.33 direct ties. This produces a GPI-RD prediction for F in

TABLE 1: Goodness of Fit[a] for Predicted and Observed Profits in Exchange Networks

Structure, Relation	Core	Rational Exchange	Equi-depend.	Expected Value	Exchange Resistance	GPI-R	GPI-RD	Observed Means
4-Line								
B-A[b]	16.0	15.0**	16.0	21.0	16.0	13.5	14.5****	14.4
t (29)	4.27	1.54	4.27	13.01	4.27	-2.54	0.18	
Stem[c]								
A-B	20.1	17.2**	18.0	22.0	18.3	14.4*	15.6****	15.9
t(13)	5.46	1.72	2.76	7.91	3.15	-1.88	-0.33	
Kite								
D-E	unstable	15.2*	12.0****	12.0****	12.5****	12.2****	13.7****	12.8
t(7)		1.95	-0.61	-0.61	-0.20	-0.46	0.77	
Double Branch 2								
F-G	16.8***	13.8	16.0****	20.2	14.6	13.3	16.3****	16.4
t(63)	0.93	-5.68	-0.82	8.38	-3.89	-6.69	-0.14	

(Column group heading: Models and Predictions)

[a] One-sample *t* tests were used to estimate the probability of no difference between prediction and observation. Larger *p* values suggest an increased likelihood that prediction and observation are identical. Degrees of freedom are in parentheses.

[b] Predictions are for profits of the first position listed in a relation; here, for example, actor B in B-A.

[c] The A-C relation is also of interest in this network. However, during the last period of the experiment, A exchanged with C only twice, both times at 14-10. This precludes meaningful comparison. With that caveat, we report only for completeness that the GPI-RD prediction of 13.7 for this relation was closest among the models, though the equidependence and rational exchange models were also close.

* $p > .05$ ** $p > .10$ *** $p > .20$ **** $p > .40$

exchange with G of 16.31. A *t* test finds no difference between this value and the observed mean, 16.36, $t(63) = -.14$, $p = .89$. While this is the closest prediction of any model, Table 1 shows that the Bienenstock and Bonacich (1992, 1993) core and Cook and Yamagishi (1992) equidependence models also make acceptable predictions for the Double Branch 2.

In sum, the GPI model incorporating resistance and degree formulations produced very accurate predictions for exchange outcomes. These predictions were superior to alternative models in their goodness of fit to experimental data: Only GPI-RD makes acceptably close predictions ($p > .40$ of no difference between predicted and observed values) for all four experimental networks.

EXPERIMENT 2: RESTRICTED-INFORMATION NETWORKS

There is a theoretical distinction between full- and restricted-information settings used for network exchange experiments. Full-information settings more closely model rational, forward-looking actors who use whatever information is available. Restricted-information settings conform better to backward-looking actors who adjust their response only on the basis of their experience in the situation. Our model requires that, minimally, three information conditions must be satisfied for negotiated social exchanges to occur:[16] (1) An actor in the network must be informed of, and have access to, other actors with whom it is possible to exchange. We assume that actors negotiate separately with each potential partner and are thus aware of each partner as a distinct person or organizational unit. Implicitly, then, actors also know the number of others with whom it is possible to exchange. (2) The actor must be informed of whether an exchange has been completed with each potential partner. (3) The actor must be informed of the magnitude of her or his profits from exchanges. In order to evaluate an offer, it must be at least ordinally scalable. This requires information on the offer's relative magnitude. In a typical experiment, this takes the form of an agreement between two partners to allocate a pool of resources at the conclusion of a given negotiation round. Implicitly, if actors know the magnitude of the offer upon which agreement was reached, then they also know the magnitude of their own shares of subsequent resource allocations stemming from the agreement. Knowledge of others' profits is not essential. Therefore, to examine the empirical scope of our model, we examine data from experiments in a new, restricted-information setting that differs in several ways from the full-information setting described above. The new setting restricts information to the minimum necessary for the operation of factors deemed important in the model.

As our theory evolves, refinements in its predictions demand that we study increasingly subtle network exchange phenomena. Consequently, our experimental setting must be made increasingly sensitive to predicted phenomena, and it must exert more stringent controls over potentially extraneous factors such as equity concerns. We have attempted to accomplish this by creating a new experimental setting that spreads the negotiation process across a larger number of rounds and limits information to the minimum necessary for negotiation and exchange. Each subject has information only on his or her own (1) negotiations and exchanges, (2) potential profit vis-à-vis particular offers received from others, and (3) realized profit when an exchange occurs. In addition, subjects negotiate for many more rounds because a subject's intra-round negotiations with a partner are simplified to a choice of three options: increase or decrease the previous offer by one "profit point" or do not change the previous offer. Cook et al. (1983) limit information in a similar fashion but allow subjects to select their offers from a wide range of possibilities on each round. The new setting then makes possible tests of our theory under information conditions similar to those of earlier restricted-information exchange experiments.

Custom software was used to configure networked PCs.[17] Subjects were isolated in separate rooms and knew only the coded designations for their own

potential exchange partners. They were informed that the shape of the full network would not be revealed and that their potential partners might have other potential partners of their own. An interactive tutorial guided subjects through the mechanics of conducting negotiations via the computerized system. On each round, subjects sent messages to a central computer telling it the lowest amount of profit they were willing to accept from each potential partner. If an agreement was reached, the subject was informed of this fact but did not know the amount of profit received by the partner; only her or his own profit was reported. Subjects completed a total of 60 rounds at the same network position. Each relation contained 30 profit points at the outset of each round, although subjects did not know the pool size. At the outset of negotiations, each partner could receive 15 points from an agreement, which was awarded as a 15-cent "bonus" for reaching an agreement. Subjects could raise or lower the amount they were willing to accept from each partner by 1 point, or leave the amount unchanged. Each 1-point change resulted in a 1-cent change in the amount of bonus for agreement. The computer declared an agreement when the sum of the profits for which two potential exchange partners were willing to settle did not exceed 30 points. If the sum was less than 30 points, the computer split the excess and awarded half to each subject in addition to the amount on which she or he had settled.

Because subjects made offers to all potential partners on each round, some could reach provisional agreements with more than one partner. Due to the fact that subjects were allowed only one agreement on each round, the central computer used the following algorithm to prioritize agreements: (1) Assign zero profit to subjects who do not reach any provisional agreement. (2) For those who remain, declare agreements for pairs of subjects whose best deals are with each other. (3) Select a subject randomly from those remaining and award her or his best deal. (4) Repeat the random selection until no more deals are possible.

This restricted-information setting differs from the full-information setting in several important ways.[18] Nevertheless, the settings are identical in several respects crucial to our model. First, in both settings the number of direct relations a subject has to others is immediately apparent and obvious. Therefore, degree can influence profit expectations. Second, in both settings, actors can over time get a sense of the range of acceptable terms through experiencing rounds in which they are excluded from exchange and rounds in which they are included. (This is true in the restricted-information setting even if some inclusions have a chance element because of computer intervention when multiple provisional exchanges could be made.) Therefore, excludability can influence expectations. On the basis of these essential similarities, we make the same predictions for profit differentials by position at equilibrium in the new restricted-information setting. That is, exchange rates in the last rounds of an experiment should be comparable across settings.[19]

RESULTS

Eleven groups participated in the Stem network. We treated the last ten agreements in a session for each relation as an indicator of its equilibrium exchange rate. This provides sufficient cases for a stringent statistical test and roughly corresponds to our use of last-period results in the full-information experiments. For the A-B relation, the last ten agreements varied by no more than a few points in all groups, allowing us to conclude that equilibrium had been reached. The maximum range over which agreements varied was 4 (M = 2.0; Std. dev. = 1.05).[20] In exchanges with B, the subject in the A position achieved mean profit of 20.13 (std. dev. = 4.29) out of a pool of 30 points, compared to the GPI-RD predicted level of 19.5. A one-sample t test found no difference between prediction and observation, $t(10) = -.49$, $p = .64$. Profit of 20.13 on a 30-point scale is equivalent to 16.10 on a 24-point scale, and thus very close to the 15.86 observed in the full-information experiment.

In previously reported experiments using the Stem network (Cook, Donnelly & Yamagishi 1992; Markovsky et al. 1993), A-C exchanges were infrequent. We had hoped that with 60 rounds in the new setting, we could establish an equilibrium value for this relation. This was not the case. A-C exchanges were still infrequent, especially during the final 30 rounds of a session. Two groups had no A-C exchanges during the last 30 rounds, and only four groups had ten or more. With such limited data, we lack confidence that equilibrium was reached. However, we attempted to test our prediction for A in the A-C relation by taking the overall mean for all A-C exchanges that occurred in the last half of a session (M = 17.46, Std. dev. = 3.77). GPI-RD predicts A will receive 17.10 profit units at equilibrium, a difference of less than half a profit unit from the observed mean. A t test found no difference between predicted and observed values, $t(82) = -.88$, $p = .38$. While this result does support the GPI-RD model, the variability in frequency of exchange argues against giving it much weight.

A significant difference was found in the C_1-C_2 relation. As with the A-B relation, we were able to use C_1's mean profit for the last ten exchanges for each group as an indicator of the equilibrium exchange rate in C_1-C_2 (M = 18.18, Std. dev. = 4.31). This difference in profit between isomorphic network positions is puzzling; all models predict an equal, 15:15, division of profit. Comparing this predicted value to the observed mean, we find that $t(10) = -2.45$, $p = .03$. The anomalous finding may be a chance occurrence, or an artifact of the experimental setting: The program treats C_1 and C_2 identically, with the one exception being that C_1 appears above C_2 as a potential partner on A's video screen. Possibly because of simple ordering, A may pay more attention to C_1 than to C_2, thereby affecting C_1's negotiations with C_2 in C_1's favor. While this might not affect the A-C or A-B equilibrium values (and thus show how robust GPI-RD predictions are for these relations), it could affect the C_1 - C_2 value. For the present, we regard this finding as a spur to additional research rather than as a disconfirmation of GPI-RD, since the finding is completely unanticipated by any alternative model.

To summarize, we replicated our test of the GPI-RD model using the Stem network in an information-restricted exchange situation. The key hypotheses were again supported.

Discussion and Conclusions

We developed a theory to explain how actors in social exchange networks reach agreements on divisions of resources. The model incorporates previous ideas about the effects that network structure has on the power of individual positions, specifically the graph-theoretical power index of network exchange theory and its weak power extension (Markovsky, Willer & Patton 1998; Markovsky et al. 1993). To this model we added theoretical ideas borrowed from several areas of sociology.

From elementary theory's concept of resistance, we borrowed the idea that actors balance two competing interests to reach agreement in exchange: (1) their "best hope" for maximum profit and (2) their "worst fear" of being excluded from exchange entirely. We combined this with an idea from network exchange theory: Likelihood of being included in exchange ranks the power of network actors. This resulted in a new assumption: Actors' best hopes and worst fears are proportional to their likelihood of being included in exchange. Actors who are frequently excluded from exchanges (and profit) are likely to lower both their maximum and minimum aspirations for profit. Conversely, actors frequently included in exchanges become accustomed to receiving profit and raise their expectations accordingly. Integrating these previous theoretical strands yielded a model that generates ratio-scale predictions for the outcomes of negotiating actors in exchange networks.

The fact that actors adjust their expectations through negotiation implies that structural power differences emerge over time. In strong power networks identified by network exchange theory, these differences never reach an equilibrium point short of the point of extreme differentiation. They continue until powerful actors receive all (or nearly all) available resources from exchange with less powerful actors. In weak power structures, an equilibrium point is reached well short of maximum differentiation. It is this equilibrium point that we attempt to predict theoretically and measure experimentally.

We felt that the equilibrium point eventually reached will likely be affected by actors' initial expectations formed on the basis of prominent features of their structural context. From network analysis we borrowed the idea that an actor's degree, the number of her or his direct ties to other actors, would influence her or his initial expectations for success in exchange. That is, actors with more direct ties would be biased toward resisting exchange offers that they would otherwise accept. We included degree as a biasing factor in predicting the equilibrium exchange point that experimental subjects would eventually reach. Results of an experimental test in a setting specifically designed to measure the equilibrium point suggest that our theoretical integration was successful. This brings up potential avenues for future research. The theory suggests that actors' expectations have significant effects on resource distribution only in weak power networks. What effects, if any, do expectations have in strong power

networks? Also, certain expectations about the social structure were shown to have important effects in weak power networks. Do other kinds of cognition have important effects? And under what conditions are cognitions likely to be more or less important?

Our extension of network exchange theory provides a number of advantages over earlier versions and current alternatives. By highlighting the ways that actors' profit expectations interact with structural properties of their locations in the network, it generates predictions for all positions in weak power networks on a ratio rather than an ordinal level of measurement. Moreover, these predictions are more accurate than those of alternative theories, and the theory generalizes across experimental designs. Our predictions closely approximate experimental results from the full-information experiments of Skvoretz and Willer (1991, 1993), the restricted-information experiments of Cook, Donnelly & Yamagishi (1992), and the results reported here for equilibrium rates in both full and restricted settings. In addition, results reported by Bienenstock and Bonacich (1993) for two weak power networks, the four-actor line and Kite, are extremely close to our predictions. Their experimental setting has features quite different from either the Skvoretz and Willer or Cook, Donnelly & Yamagishi designs. This remarkable convergence of experimental results in different settings demonstrates both the increased precision of the theory and its enhanced generality.

The empirical results also clearly suggest that equity concerns are not inextricably woven into social exchange network settings. This is not to say that equity effects are unimportant, but rather that equity is a distinct process that may or may not be activated in a given social context and that depends upon whether certain conditions are satisfied (Markovsky 1985). In developing our restricted-information setting, we struggled with the powerful effects of subjects' equity concerns when they felt they were receiving less than a partner who in other ways was their status equal. In some cases, subjects would refuse to exchange in as many as 50 out of 60 rounds because another subject would receive more profit than they would. That is, subjects would refuse five or six dollars in pay to avoid receiving a few pennies less than their partner in exchange. This study demonstrates that once equity concerns are controlled, different experimental settings produce comparable structural effects on resource distributions resulting from exchange. Structural positions have an effect on power independent of equity concerns. An interesting area for further inquiry is exactly how equity effects combine with the effect of structural position under different conditions to produce power and profit differences in social exchange networks.

Let us note that the equation for actor profit, P_i, converges with part of Cook and Yamagishi's (1992) theory — their equation $R_{ij} = (P_{ij} + A_{ij} - A_{ji})/2$. In this model, R_{ij} is the profit that actor i obtains in exchange with partner j and corresponds to our P_i; P_{ij} is the pool size and corresponds to P; A_{ij} is the best alternative available to i, which corresponds to C_{ij}, i's expected conflict outcome. That is, i's best alternative is the least amount of profit i expects if exchange with j does not occur. A_{ji} is j's best alternative, i.e., C_{ji}. Despite these similarities, however, our model diverges from Cook and Yamagishi's in significant ways. Unlike A_{ij}, which refers to the objective profit under "conflict," C_{ij} is assumed

to be a subjective assessment or expectation of long-range profit from failure to reach agreement. We believe that one reason for our model's predictive success stems from this incorporation of the actor's point of view. This allows the new model to generate contrasting predictions that are here shown to be significantly more accurate than alternatives.

Although the predictions that we derived are accurate for the networks tested, these findings tell us only that the model is developing in potentially fruitful directions. Establishing its broader generality will require continued testing in a wider variety of networks. Further enhancements will be required to allow predictions with theoretical restrictions further relaxed. Of course, these have been our goals all along: to generate increasingly precise and accurate predictions for network exchange outcomes under increasingly robust conditions. With the theoretical and empirical developments reported above, we have worked toward achieving these goals.

APPENDIX: Network Exchange Theory

Key Terms

p = power	h = length of the longest nonintersecting path
d = domain	m_{idk} = number of i's nonintersecting paths of length k in domain d.
i, j = network positions	e = maximum exchanges per round

A network path is a series of connected positions, e.g., A-B-C-D. Two paths stemming from a given position i are nonintersecting only if i is the only position common to both. Thus, relative to the A position, A-B-C-D and A-B-E-F are intersecting paths (at B), but neither intersects with A-G-H.

Domains are network substructures. When $e > 1$, a position can have different power indices within different domains (see Markovsky, Willer & Patton 1988). The following criterion determines whether i and j are in the same domain. First, define an $e+$ position as one having more than e relations. Then, given the set V of all positions on a path (i.e., a series of linked positions) between i and j, i and j are in the same domain only if there exists a path such that either V = {Ø}, or all positions in V are $e+$ positions.

SCOPE CONDITIONS

The theory is deemed applicable when the following conditions are met or approximated:

1. All actors use identical strategies in negotiating exchanges.
2. Actors consistently excluded from exchanges raise their offers.
3. Those consistently included in exchanges lower their offers.
4. Actors accept the best offer they receive and choose randomly in deciding among tied best offers.
5. Each position is related to, and seeks exchange with, one or more other positions.
6. At the start of an exchange round, equal pools of positively valued resource units are available in every relation.
7. Two positions receive resources from their common pool if and only if they exchange (Markovsky, Willer & Patton 1988:223).

AXIOMS

Four axioms determine the relative power of connected positions and whether they will exchange:

Axiom 1. $$P_{id}(e_d) = \left(\frac{1}{e_d}\right)\sum_{k=1}^{h}(-1)^{(k-1)}m_{idk}$$

Axiom 2. i seeks exchange with j only if $p_i > p_j$, or if $(p_i - p_j) \geq (p_i - p_k)$ for all k related to i.

Axiom 3. i and j can exchange only if each seeks exchange with the other.

Axiom 4. If i and j exchange, then i receives more resources than j if and only if $p_i > p_j$.

APPENDIX: Network Exchange Theory

WEAK AND STRONG POWER

Markovsky et al. (1993) provide an iterative method, GPI^2, used to derive more precise hypotheses.

Step 1. Apply Axiom 2 to determine exchange seeks.

Step 2. Apply Axiom 3 to remove nonmutual exchange seeks.

Step 3. Apply Axiom 1 to each subnetwork that results from Step 2; $p = 1$ for isolates.

Step 4. Re-form the full network with new p values; repeat from Step 1 until values do not change.

If after the above analyses $p_i > p_j$ for actors in the i - j relation, then i has strong power over j, and it is predicted that profits from exchange in this relation will favor i and approach their maximum differentiation. If $p_i = p_j$, then there are exactly two possibilities: either the actors are equal in power, or else one has weak power over the other.

Weak power differences are detected using a probability analysis. The analysis assumes that actors seek exchanges randomly among their potential partners and counts relative proportions of mutual exchange seeks as exchange likelihoods. If i and j have unequal exchange likelihoods, then the actor with the higher likelihood is predicted to have a weak power advantage over the other. This is identified as GPI_3. Examples of its application are provided in Markovsky et al. (1993) and in the theoretical discussion of this article. A computer program for calculating likelihood of inclusion is available at no charge from the second author.

Notes

1. It should be noted, however, that network exchange theory predictions also generally agree with data from exchange situations with greater information restrictions, e.g., Cook, Donnelly, and Yamagishi (1992), Cook and Emerson (1978), and Cook et al. (1983), and also from the setting used by Bienenstock and Bonacich (1993).

2. These conditions are probably sufficient but not necessary to determine whether a network displays strong power. In general, casual inspection often fails to classify networks properly as strong power or weak power. Full application of the GPI method is required.

3. Theoretical integration requires integration of notation systems as well. Markovsky et al. (1993) use the notation $p\{i\}$ to denote the probability of inclusion of position i in an exchange network. But the letter p also occurs in resistance equations to denote profit. To avoid confusion and simplify our notation, we switch to l_i (actor i's likelihood of exchange).

4. C_i is similar to Thibaut and Kelley's (1959) "comparison level for alternatives" or CL_{ALT}, i.e., "the lowest level of outcomes a member will accept in the light of available alternative opportunities" (21). M_i and C_i define the range of possible offers. The model does not assume that actors have objective knowledge of their values. "Best hopes" and "worst fears" need not be reasonable, though actors are likely to refine their estimates as they interact. We have again simplified the notation of earlier presentations of the theory: e.g., Willer, Markovsky & Patton (1989) use $P_{MAX}(A)$ to represent M_i and $P_{CON}(A)$ to represent C_i.

5. Assumptions are evaluated on their effectiveness in producing testable hypotheses that conform well with observation. The assumption that actors exchange when their resistances are equal has been very fruitful in previous studies (Skvoretz & Willer 1993; Willer 1987), including cross-national comparisons (Willer & Szmatka 1993).

6. Cook and Yamagishi (1992) also suggest that the idea of a limit to power use in networks holds promise for a general formula to predict resource distribution. Willer (1987) demonstrat-

ed such generality when he applied the resistance model to a wide variety of network situations both inside and outside the laboratory.

7. Individuals participating in experiments or acting in natural exchange situations will exhibit a range of "best hopes" and "worst fears." This in no way interferes with the model's ability to predict exchange rates. Coalitions among actors are ruled out by the scope conditions of the theory though they may occur often in exchange situations. Erger (1993) has extended the theory to include the effects of coalitions.

8. Markovsky et al. (1993) provide support for this idea. They found that ordinal predictions for weak power networks based on likelihood of being included were more strongly corroborated for experienced than for inexperienced subjects.

9. While simple proportionality is a straightforward way to incorporate likelihood of being included into the resistance model, other specifications are possible. For example, Skvoretz and Willer (1993) take the difference between M_i and P_i, then raise it to the power of l_i. Our model is the simplest expression we could devise of the theoretical idea that actors' worst fears and best hopes in the exchange situation depend on — and are proportional to — the likelihood of their being included in exchanges.

10. The mathematical derivation is available on request from the first author.

11. Markovsky (1988) specifies the conditions under which anchoring will occur: judgments are indeterminate, an anchor is available, and anchors are salient. These conditions are satisfied in experimental tests of network exchange theory. Markovsky's "anchoring proposition" predicts when assimilation as opposed to contrast effects will be observed. According to this proposition, assimilation would be predicted in the present context because degree informs best-hope and worst-fear outcomes, each of which appears on the same scale as the "response" variable, i.e., expected profit. (An anchor on the stimulus scale — as in the temperature example — produces a contrast effect.)

12. An actor's maximum expectation for profit may differ among exchange partners in the same way.

13. In a replication using a different experimental exchange setting, Bienenstock and Bonacich (1993) obtained similar results.

14. Some readers of earlier versions of this article noted the similarity between network exchange and noncooperative game theory (e.g., Harsanyi 1980; Nash 1951; Osborne 1990; Rosenthal & Rubinstein 1984; Rubinstein 1982, 1991). Also, the few experimental tests of noncooperative game theory use experimental situations similar to that used in network exchange experiments but without the complication of network structure (see, e.g., Nydegger & Owen 1974). While intriguing, these similarities mask very real difficulties in applying noncooperative game theory to network exchange. Rubinstein (1982) states the bargaining problem in noncooperative game theory as "Two individuals have before them several possible contractual agreements. Both have interests in reaching agreement but their interests are not entirely identical. What 'will be' the agreed contract, assuming that both parties behave rationally?" (97). He goes on to distinguish this problem from two others: "(i) the positive question — what is the agreement reached in practice; (ii) the normative question — what is the just agreement." Perhaps because of these distinctions, noncooperative game theory places little emphasis on theory testing through experimental or field research and does not fare well in experimental tests. Network exchange theories place more emphasis on the "positive question," on how subjects behave in controlled settings. Experimental results are then used to inform theoretical development in cumulative research programs. Bienenstock and Bonacich have made the most successful use of game theory to analyze network exchange structures.

15. Markovsky et al. (1993) and Skvoretz and Willer (1993) analyze data from these full information experiments. Markovsky et al. (1993) use data from the Stem and Kite networks; Skvoretz and Willer use data from all four experiments. Their analyses are based on all rounds of the experiments. Here we use data from just the last period, which is four rounds long. Although suitable for testing ordinal predictions, using the mean of all rounds in an experiment as an indicator of power is problematic for testing exact predictions. For example, exchange may begin at an even split of the profit pool, 12:12, in early rounds then progress to a stable pattern of 20:4 exchanges. In this case, 20:4 is a good estimate of the power difference

in the relationship. The mean exchange rate for all rounds (about 16:8) would seriously underestimate the magnitude of the equilibrated power difference.

16. Our model requires these assumptions; we do not assume that all naturally occurring social exchanges satisfy these conditions.

17. This system was designed to be relatively "low-tech" and portable to other laboratories. The software is written in Microsoft QuickBASIC (4.5), and PCs are connected in a ring configuration via cables connected to standard serial ports. The ring consists of one master control PC and any number of subject PCs. The program is available from the authors upon request.

18. Equity concerns, for instance, are controlled in the two settings in different ways. If actors feel the exchange situation is unfair, they may refuse to accept the best offer available to them. The full-information setting described solves the potential equity problem by rotating subjects through all positions. Actors disadvantaged in one position know they will be compensated when they rotate through an advantaged position. Restricted-information settings in which subjects typically do not change positions solve the problem by not telling a subject the earnings of his or her partner in order to prevent comparison of subject's rewards with partner's rewards.

19. Because of the differences in intraround negotiation options and total number of rounds between the two settings, we would not expect averages from all rounds to be similar across settings. The restriction to equilibrium rates is essential to the "no setting difference" prediction.

20. In contrast, the first ten A-B agreements for each group varied more widely; the maximum range was 9 ($M = 4.23$, Std. dev. = 2.35). The mean range of the first ten agreements was significantly greater than the mean range of the last ten agreements, $t(10) = 3.09$, $p = .01$.

References

Bienenstock, Elisa Jayne, and Phillip Bonacich. 1992. "The Core as a Solution to Exclusionary Networks." *Social Networks* 14:231-43.

_____. 1993. "Game Theory Models for Exchange Networks: Experimental Results." *Sociological Perspectives* 36:117-35.

Brennan, John S. 1981. "Some Experimental Structures." Pp. 188-204 in *Networks, Exchange and Coercion: The Elementary Theory and Its Application,* edited by David Willer and Bo Anderson. Elsevier.

Blau, Peter M. 1964. *Exchange and Power in Social Life.* Wiley.

Coleman, James S. 1973. *The Mathematics of Collective Action.* Aldine.

_____. 1990. *Foundations of Social Theory.* Harvard University Press.

Cook, Karen S., Shawn Donnelly, and Toshio Yamagishi. 1992. "The Effect of Latent Paths on Power in Exchange Network Structures." Paper presented to the annual meeting of the American Sociological Association, August, Pittsburgh, PA.

Cook, Karen S., and Richard M. Emerson. 1978. "Power, Equity and Commitment in Exchange Networks." *American Sociological Review* 43:721-39.

Cook, Karen S., Richard M. Emerson, Mary R. Gillmore, and Toshio Yamagishi. 1983. "The Distribution of Power in Exchange Networks: Theory and Experimental Results." *American Journal of Sociology* 89:275-305.

Cook, Karen S., Mary R. Gillmore, and Toshio Yamagishi. 1986. "Point and Line Vulnerability as a Basis for Predicting the Distribution of Power in Exchange Networks." *American Journal of Sociology* 92:445-48.

Cook, Karen S., and Toshio Yamagishi. 1992. "Power in Exchange Networks: A Power-Dependence Formulation." *Social Networks* 14:245-65.

Emerson, Richard M. 1962. "Power-Dependence Relations." *American Sociological Review* 27:282-98.

____. 1972. "Exchange Theory Part II: Exchange Relations and Network Structures." Pp. 58-87 in *Sociological Theories in Progress*. Vol. 2, edited by Joseph Berger, Morris Zelditch Jr., and Bo Anderson. Houghton Mifflin.

Erger, Jeff. 1993. "Structural Cooperation in Exchange Networks." Paper presented at the annual meeting of the American Sociological Association, Miami Beach, August.

Friedkin, Noah. 1992. "An Expected Value Model of Social Power: Predictions for Selected Exchange Networks." *Social Networks* 14:213-29.

Harsanyi, J.C. 1980. "Analysis of a Family of Two-Person Bargaining Games with Incomplete Information." *International Journal of Game Theory* 9:65-89.

Heckathorn, Douglas D. 1983. "Extensions of Power-Dependence Theory: The Concept of Resistance." *Social Forces* 61:1206-31.

Helson, Harry, and A. Kozaki. 1968. "Anchor Effects Using Numerical Estimates of Simple Dot Patterns." *Perception and Psychophysics* 4:163-64.

Homans, George C. 1958. "Social Behavior as Exchange." *American Journal of Sociology* 63:597-606.

____. 1974. *Social Behavior: Its Elementary Forms*. 2d ed. Harcourt Brace Jovanovich.

Kahneman, Daniel, Paul Slovic, and Amos Tversky. 1982. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press.

Markovsky, Barry. 1985. "Toward a Multilevel Distributive Justice Theory." *American Sociological Review* 50:822-39.

____. 1988. "Anchoring Justice." *Social Psychology Quarterly* 51:213-24.

____. 1992. "Network Exchange Outcomes: Limits of Predictability." *Social Networks* 14:267-86.

Markovsky, Barry, John Skvoretz, David Willer, Michael Lovaglia, and Jeffrey Erger. 1993. "The Seeds of Weak Power: An Extension of Network Exchange Theory." *American Sociological Review* 58:197-209.

Markovsky, Barry, David Willer, and Travis Patton. 1988. "Power Relations in Exchange Networks." *American Sociological Review* 53:220-36.

____. 1990. "Theory, Evidence and Intuition." (Reply to Yamagishi and Cook). *American Sociological Review* 55:300-305.

Marsden, Peter V. 1983. "Restricted Access in Networks and Models of Power." *American Journal of Sociology* 88:686-717.

Nash, John F. 1950. "The Bargaining Problem." *Econometrica* 18:155-62.

____. 1951. "Non-Cooperative Games." *Annals of Mathematics* 54:286-95.

____. 1953. "Two-Person Cooperative Games." *Econometrica* 21:129-40.

Nydegger, R.V., and G. Owen. 1974. "Two-Person Bargaining: An Experimental Test of the Nash Axioms." *International Journal of Game Theory* 3:239-49.

Osborne, Martin J. 1990. "Signaling, Forward Induction, and Stability in Finitely Repeated Games." *Journal of Economic Theory* 50:22-36.

Rosenthal, R.W., and A. Rubinstein. 1984. "Repeated Two-Player Games with Ruin." *International Journal of Game Theory* 13:155-77.

Rubinstein, Ariel. 1982. "Perfect Equilibrium in a Bargaining Model." *Econometrica* 50:97-109.

____. 1991. "Comments on the Interpretation of Game Theory." *Econometrica* 59:909-24.

Skvoretz, John, and Thomas J. Fararo. 1992. "Power and Network Exchange: An Essay toward Theoretical Unification." *Social Networks* 14:325-44.

Skvoretz, John, and Michael J. Lovaglia. N.d. "Who Exchanges with Whom: Structural Determinants of Exchange Frequency in Negotiated Exchange Networks." *Social Psychology Quarterly*. Forthcoming.

Skvoretz, John, and David Willer. 1991. "Power in Exchange Networks: Setting and Structural Variations." *Social Psychology Quarterly* 54:224-38.

———. 1993. "Exclusion and Power: A Test of Four Theories of Power in Exchange Networks." *American Sociological Review* 58:801-18.

Thibaut, John W., and Harold H. Kelley. 1959. *The Social Psychology of Groups*. Wiley.

Willer, David. 1981. "Quantity and Network Structure." Pp. 108-27 in *Networks, Exchange and Coercion: The Elementary Theory and Its Application*, edited by David Willer and Bo Anderson. Elsevier.

———. 1986. "Vulnerability and the Location of Power Positions." *American Journal of Sociology* 92:441-44.

———. 1987. *Theory and the Experimental Investigation of Social Structures*. Gordon and Breach.

Willer, David, and Barry Markovsky. 1993. "Elementary Theory: Its Development and Research Program." Pp. 323-63 in *Theoretical Research Programs: Studies in the Growth of Theory*, edited by Joseph Berger and Morris Zelditch Jr. Stanford University Press.

Willer, David, Barry Markovsky, and Travis Patton. 1989. "The Experimental Investigation of Social Structures." Pp. 313-53 in *Sociological Theories in Progress: New Formulations*, edited by Joseph Berger and Morris Zelditch Jr. Sage.

Willer, David, and Travis Patton. 1987. "The Development of Network Exchange Theory." Pp. 199-242 in *Advances in Group Processes*. Vol. 4, edited by Edward J. Lawler and Barry Markovsky. JAI.

Willer, David, and Jacek Szmatka. 1993. "Cross National Experimental Investigation of Elementary Theory: Implications for the Generality of the Theory and the Autonomy of Social Structures." Pp. 37-81 in *Advances in Group Processes*. Vol. 10, edited by Edward J. Lawler, Barry Markovsky, Karen Heimer, and Jodi O'Brien. JAI.

Yamagishi, Toshio, and Karen S. Cook. 1990. "Power Relations in Exchange Networks: A Comment on 'Network Exchange Theory.'" *American Sociological Review* 55:297-300.

Persistence and Change in the Protestant Establishment, 1930-1992*

JAMES D. DAVIDSON, *Purdue University*
RALPH E. PYLE, *University of Nevada, Reno*
DAVID V. REYES, *Purdue University*

Abstract

Three Protestant denominations comprising "the Protestant Establishment" (Episcopalian, Presbyterian, and Congregationalist/ United Church of Christ) played a major role in shaping the American way of life. While some scholars claim that Protestant hegemony persists, others argue that we live in a post-Protestant era. We assess these conflicting views by examining the religious affiliations of persons listed in Who's Who in America *in 1930 and 1992. The results indicate that the Protestant Establishment has lost some ground since the 1930s but remains overrepresented among the nation's power and cultural elites. Among other groups, Jews have gained the most. Catholics also have gained, but remain slightly underrepresented among both cultural and power elites. The status of other religious groups is also discussed.*

Three Protestant denominations (Episcopalian, Presbyterian, and Congregationalist [since 1957 United Church of Christ]) have profoundly influenced the social and cultural fabric of American society. These groups emerged as social, economic, and political elites during the first century of American history (Baltzell 1958, 1964; Gaustad 1968; Hutchison 1989). They created an American way of life that reflected their worldviews and worked to their advantage socially, economically, and politically. Because of their prominence among America's elite, they became known as "the Protestant Establishment" (Baltzell 1964).[1]

The Protestant Establishment became so highly institutionalized that some scholars believe its influence upon America's core culture can still be felt (Feagin & Feagin 1993; Greeley 1977; Rossides 1990). While granting that white Anglo-Saxon Protestants (WASPs) are a declining percentage of the total population and that non-WASPs have made some inroads into the upper

*** This research was supported in part by Grant #93-0022 from the Louisville Institute for the Study of Protestantism and American Culture. This article was presented at the annual meeting of the Association for the Sociology of Religion, Los Angeles, August 1994.** ***Please direct correspondence to James D. Davidson, Department of Sociology and Anthropology, 1365 Stone Hall, West Lafayette, IN 47907-1365.***

 Social Forces, September 1995, 74(1):157-175

echelons of American society, these observers emphasize the tenacity of the Protestant Establishment. They contend that an institutionalized system of Establishment-oriented policies and practices persists in most spheres of American life (especially in business and politics), and that the percentage of Episcopalians, Presbyterians, and members of the United Church of Christ (UCC) among the elite is not significantly different from what it was earlier in this century.

However, other scholars argue that major societal changes have eroded the Protestant Establishment's influence (Baltzell 1976, 1991; Christopher 1989; Hammond 1992; Hutchison 1989; Schrag 1970). These writers believe that as a result of forces such as the increased religious pluralism at the turn of the century, the G.I. Bill at the end of World War II, the civil-rights and black power movements of the 1950s and 1960s, and greater cultural pluralism during the 1960s and 1970s, access to positions of social, economic, and political dominance is no longer limited to the Protestant Establishment. They claim that other religious groups (especially Catholics and Jews) have not only achieved socioeconomic parity with Protestants, but they also have penetrated the highest echelons of power and privilege.

Thus, there are sharp differences of opinion on the tenacity or decline of the Protestant Establishment. In an attempt to see what actually happened to the Protestant Establishment between the 1930s and the 1970s, Davidson (1994) recently compared Fry's 1930-31 data from *Who's Who in America* (Fry 1933a, 1933b) with Verba and Orren's 1976-77 study of elites (Verba & Orren 1985). Davidson's results emphasized the persistence of the Protestant Establishment (with Episcopalians, Presbyterians, and UCCs continuing to be overrepresented among the elite), though some groups, especially Jews and Catholics, had made important strides. Davidson also argued that as a result of economic and political developments in the 1980s (e.g., the tendency of the Ronald Reagan and George Bush administrations to appoint WASPs to high political office), the disproportionate influence of the Protestant Establishment probably would persist well into the future. However, Davidson argued that we need more direct longitudinal comparisons of *Who's Who* in the 1930s and 1990s if we are to understand the extent to which the Protestant Establishment persists or has been replaced by a more religiously pluralistic elite. We have just completed such an analysis.

Theoretical Orientation

Building on our earlier work (Davidson 1985, 1994), we use a neo-Weberian, interest group perspective to frame our analysis. This "modified fair shares" perspective (Davidson 1985, 1994; Ryan 1981) stresses the conflict between religious groups. Using policies and practices that we describe below, Episcopalians, Presbyterians, and Congregationalists gained the upper hand over other religious groups and shaped social institutions according to their values and interests. The institutionalization of policies and practices working to WASPs' advantage and the disadvantage of other groups contributed to the persistence of religious inequality, especially in spheres most directly affecting the creation

and distribution of resources (i.e., business and politics). This perspective also recognizes that other religious groups oppose these obstacles, trying to remove or circumvent structural biases against them and to improve their lot in life. As we indicate below, these efforts produce social changes and mobility that sometimes lead to greater access to power and privilege. In the next two sections, we use this theoretical perspective to specify some of the reasons that the Protestant Establishment persists and why it has changed.

WHY THE PROTESTANT ESTABLISHMENT PERSISTS

The Protestant Establishment's prominence in the economic and political spheres became especially important, and it was overrepresented and very influential in most cultural spheres as well (Baltzell 1958, 1964; Feagin & Feagin 1993; Konolige & Konolige 1978). So great was its prominence that writers have used phrases like "the Episcopalianization of the whole upper class" to describe its influence (Baltzell 1958:391).

From this lofty position, the Protestant Establishment was able to shape the nation's public institutions and cultural orientations. Mainline Protestantism's religious and moral perspectives were taken for granted in the nation's public schools (Beale 1938; Lichliter 1946; Moberg 1962; Sweet 1948). Its worldview (the "Protestant ethic") also was built into the nation's economy, especially its emphasis on work and austere attitudes toward people who do not work and are poor (Kluegel & Smith 1986; Tropman 1986). This Anglo-Saxon core culture made it relatively easy for mainline Protestants to achieve and maintain elite status while also making it difficult for other groups to make it to the top (Gordon 1964).

In the process, religious prejudice and discrimination became an important part of American history. In the eighteenth and nineteenth centuries Protestant hostility toward Catholics reduced levels of Irish immigration, prevented Catholics from owning land and voting, and led to high levels of Catholic conversion to Protestantism (Anderson 1970; Clark 1973; Feagin & Feagin 1993; Gaustad 1968; O'Brien 1919). Religious prejudice became even more pronounced during the mid and late 1800s, when groups such as the Know-Nothings and the KKK fought to protect the nation's Protestant heritage against Catholics and Jews (Anderson 1970; Feagin & Feagin 1993). The Immigration Act of 1924 favored northern Europeans while limiting immigration from eastern and southern Europe (Anderson 1970; Feagin & Feagin 1993), where so many Catholics and Jews were coming from. Newspapers and other writings of the time were full of articles extolling the virtues of Anglo-Saxon culture and depicting non-WASPs as intellectually and morally inferior (Curtis 1971; Feagin & Feagin 1993).

But WASP hegemony is not just a thing of the distant past. WASP neighborhoods (e.g., sections of Philadelphia's "Main Line"), prep schools, country day schools, and social clubs persist as major threads in America's social fabric (Baltzell 1958, 1964; Domhoff 1967, 1970, 1983). During the suburbanization era following World War II, upper-income WASP communities still used "gentlemen's agreements," which made it difficult for non-WASPs to obtain housing (*Time* 1960). When President Lyndon Johnson announced changes in the nation's

immigration laws in 1965 (Feagin & Feagin 1993), the influx of new ethnic groups, who were willing to work for minimum wage or less, contributed to the prosperity of established elites and increased tensions among the nation's racial and ethnic minorities. The political conservatism of the 1970s and 1980s, especially evident during the Reagan and Bush administrations, reaffirmed the legitimacy of the WASP way of life ("back to the basics") and reversed many of the gains that racial, ethnic, and religious minorities had made in the 1960s (Feagin & Feagin 1993; Hurst 1992; Phillips 1990). Presbyterian Ronald Reagan was followed by Episcopalian George Bush, and both men tended to select other WASPs for high-ranking appointments with the expectation that these political leaders would create social policies consistent with the Protestant Establishment's core culture.

As these examples suggest, the Protestant Establishment uses several mechanisms to perpetuate its prominence. One mechanism is inheritance. Tax laws encourage elites to pass their wealth along to their sons and daughters (Phillips 1990). That is why every year when *Forbes* magazine lists the nation's 400 wealthiest people, it finds that the single biggest source of wealth is inheritance. Approximately 40% of the nation's wealthiest people have inherited their money from their families (Hurst 1992).

Another mechanism is legacy admissions at elite schools. With their enormous wealth, members of the Protestant Establishment have convinced prep schools and prestigious colleges and universities that it is in their financial best interest to set aside a certain number of positions for their sons and daughters (Karabel 1984; Karen 1991; Levine 1980). WASP elites have been especially adept at getting their children into the most prestigious New England boarding schools and Ivy League colleges. Schools such as Groton, Choate, and Taft are widely known to be steppingstones for WASP children who aspire for admission to Harvard, Princeton, and Yale (Armstrong 1990; Cookson & Persell 1985; Domhoff 1983; Dye 1983; Hurst 1992; Zweigenhaft & Domhoff 1991).

A third mechanism is appointment to high office. Members of the Protestant Establishment often select other Episcopalians, Presbyterians, and UCCs to lead their companies and political administrations (Domhoff 1983). George Bush, an Episcopalian, selected James Baker, another Episcopalian, to be his secretary of state. Then he chose David Souter, another Episcopalian, to be his first nominee to the Supreme Court. Later, he chose Clarence Thomas for the Supreme Court. Thomas was married to an Episcopalian and, though raised Catholic himself, was an active member of an Episcopal church in Virginia at the time of his appointment.

Stressing a status persistence viewpoint,[2] we believe that the Protestant Establishment remains largely intact, with Episcopalians, Presbyterians, and UCCs continuing to be overrepresented among the nation's elite, especially in the economic and political spheres, which have the most direct impact on the creation and distribution of scarce resources (Alba & Moore 1982; Davidson 1994; Domhoff 1983; Dye 1983).

CHANGE

But we also expect at least some decline in Protestant hegemony. There has been some movement away from cultural norms promoting a Protestant Establishment (Carroll & Roof 1993; Hammond 1992; Reed 1988). Ethnocentric religious barriers that previously limited the participation of non-WASPs in important economic, political, and cultural spheres have been challenged, and in some cases removed (Baltzell 1991; Christopher 1989). Since the 1960s, the nation has articulated an alternative to the Anglo conformity model that fostered an emphasis on assimilation (Gordon 1964). The increasing emphasis on multiculturalism is an expression of a growing acceptance of ethnic and religious pluralism. In this respect, the nation is more universalistic, and less religiously particularistic, than it was in previous eras, which enables formerly marginalized groups to move into prestigious positions once reserved for members of the Protestant Establishment. We have had Catholic and Baptist presidents, though Jews have not yet attained the Oval Office.

Many sociologists of religion (Carroll & Roof 1993; Finke & Stark 1992; Hammond 1992; Johnson 1985) argue that the decline in the size of the most prestigious mainline Protestant denominations indicates the wavering influence of an older Protestant Establishment. A number of analysts also suggest that denominations are no longer as important as they once were (Carroll & Roof 1993; Christopher 1989; Hammond 1992; Wuthnow 1988). Wuthnow, for example, argues that educational levels, rather than religious affiliation, separate people into distinct status categories. Declining Protestant-Catholic-Jewish tensions after World War II and increasing ecumenism and interfaith contact also imply a lessening of religious divisions (Herberg 1960; Roof & McKinney 1987). More frequent religious intermarriage indicates that traditional religious boundaries have lost some of their impact on patterns of social interaction (Kalmijn 1991).

Social mobility among non-Protestants reinforces the claim that religious distinctions are on the decline. In the last 60 years, Catholics and Jews have made impressive gains in their socioeconomic standing. Jews have made significant progress in attaining positions of prominence after World War II, especially in intellectual and cultural spheres (Alba & Moore 1982; Lieberson & Carter 1979; Weyl 1966). Descendants of East European Jews moved from the ghetto to the suburb in two generations (Baltzell 1964; Herberg 1960; Sklare & Greenblum 1979). Catholics also substantially improved their socioeconomic position, having met or surpassed Protestants on most measures of economic status (D'Antonio et al. 1989; Gallup & Castelli 1987; Glenn & Hyland 1967; Greeley 1977; Marger 1991; Roof 1979). Non-Protestants have gained a foothold in important cabinet and political posts (Mintz 1975), and they are increasingly represented on prestigious policy commissions and cultural boards (Christopher 1989; Zweigenhaft & Domhoff 1982).

Some of the gains made by non-WASPs can be attributed to the decline of discriminatory measures. Civil-rights legislation of the 1960s outlawed job discrimination on the basis of ascribed characteristics such as race or religious background (Alba 1990; Marger 1991). Formal religious quotas at elite eastern colleges have been removed, as have restrictive policies at the most exclusive

metropolitan social clubs (Baltzell 1986). Affirmative action policies have worked to broaden the ethnic composition of those in professional spheres (Zweigenhaft & Domhoff 1991). These and other measures have removed impediments that prevented the advance of non-WASPs.

Gains made by non-WASPs also result from changes in the structure of education. To a large degree, education has served as the enabling mechanism propelling Catholics and Jews into positions of influence during the last fifty years. The G.I. Bill of 1944, for example, broke down ethnic barriers to college opportunity and gave a boost to the descendants of immigrants who otherwise would not have had the opportunity to attend college (Christopher 1989; Keller 1991). Higher education enrollments in 1985 were more than eight times what they were in 1940 (Kingston & Lewis 1990). Recently, there has been a decline in the proportion of Eastern prep school graduates admitted to Ivy League schools (Cookson & Persell 1985; Zweigenhaft & Domhoff 1991), thus limiting the influence of the most prestigious Protestant boarding schools.

Advances made by non-Protestant groups also result from structural changes in the economy. Some analysts argue that economic expansion after World War II and the attendant increase in bureaucratization and scientific specialization have led to greater professional opportunities for capable individuals from diverse backgrounds (Keller 1991; Reed 1988). The movement from an emphasis on ascription to an achievement orientation favors talented ethnic groups over old family descendants (Jaher 1980). Baltzell (1986) suggests that after World War II we witnessed a change from a prewar social structure dominated by a Protestant elite to a classless bureaucracy rewarding educational achievement. As a consequence, non-Protestants have increasingly moved into prestigious educational, legal, and medical positions, although their representation as corporate officers remains limited (Alba & Moore 1982; Sturdivant & Adler 1976).

The cumulative effect of these processes is that non-Establishment groups, especially Catholics and Jews, have made significant progress in their overall socioeconomic status. This progress does not guarantee access to elite status. However, increasing percentages of Catholics and Jews probably have attained elite positions, especially in cultural fields, such as education or arts and letters, less so in business and politics, where we think the Protestant Establishment continues to be most overrepresented.

Methods

This study analyzes the religious affiliations of individuals listed in *Who's Who in America* in 1930-31 and 1992-93. *Who's Who* is probably the best known, and most highly regarded, listing of persons who have achieved special prominence or distinction in their chosen fields. It is probably the best source of information on the religious preferences of elite individuals, for whom survey data are rare.

Two classes of individuals are included in *Who's Who*: persons of significant achievement in various lines of endeavor and those included because of the official positions they occupy.[3] There is some degree of overlap between the categories. *Who's Who* editors acknowledge that most persons included because

of their official positions would also be admitted because of their achievements (Marquis Who's Who 1992).

There has been a gradual increase in the proportion of the American population listed in *Who's Who,* rising from 2.40 names per 10,000 in 1930 to 3.21 names per 10,000 in 1992. However, *Who's Who* editors insist that there has been a consistent application of admission criteria over the years and explain periodic fluctuations in the inclusion rate as a result of expansion and contraction of occupational fields (Larson 1958). The criterion of "reference interest" has been applied consistently over the years. While *Who's Who* is not without its limitations (Lieberson & Carter 1979; Pyle 1995), this roster of eminent Americans serves as a useful tool for studying changes in the denominational preferences of elite Americans from the 1930s to the 1990s.

A number of other studies have analyzed the religious preferences of *Who's Who* listees (Ament 1927; Davis 1953; Fry 1933a, 1933b; Huntington & Whitney 1927; Kirk 1957; Newcomer 1955; Selth 1987; Weber 1933). Of all these, Fry's is the most complete. Fry tabulated religious and occupational information on everyone listed in the 1930-31 edition of *Who's Who* (*N* = 29,623).

We selected a 1-in-20 systematic sample of the approximately 80,400 names listed in the 1992-93 edition. Our large *N* of 4,018 also minimizes sampling-based measurement error and provides the best available basis for estimating the contemporary status of the Protestant Establishment.[4]

Fry's 1930-31 analysis indicated that 56% of the entries reported a denominational preference. In our 1992-93 sample, 34% listed one. While some of this difference might be related to an increase in the percentage of the U.S. population reporting no religious preference (Glenn 1987; Roof & McKinney 1987), the difference probably has more to do with an increased tendency for elites to withhold personal information in public documents such as *Who's Who.* The decline in religious reporting is consistent with a reduction in the reporting rate for other personal data, such as listees' home addresses, club memberships, and political affiliations (Priest 1982; Pyle 1995; Selth 1987).

To explore the status of religious groups in different spheres, we put *Who's Who* listees in two broad occupational categories, power elites (occupations with considerable control over the creation and distribution of economic and political resources) and cultural elites (fields with more influence over ideas and cultural symbols). Power elites included bankers, businessmen, politicians, diplomats, judges, lawyers, and military officers (32% of all elites in 1930 [*N* = 9,314]; 48% in 1992 [*N* = 1,884]). Cultural elites included educators, scientists, doctors, engineers, editors and authors, artists and actors, and religious and social workers (68% of all elites in 1930 [*N* = 19,698]; 52% in 1992 [*N* = 2,049]). Those not falling within either of these divisions were left out of the occupational analysis (2.1% of the original sample in both years).

Z scores are used to indicate significant differences in the religious affiliations of *Who's Who* entries in 1930 and 1992. Since these comparisons do not take into account changes in denominational market shares among the general population over time (Finke & Stark 1992; Roozen 1993), ratios also are used to indicate denominational affiliation in *Who's Who* relative to denominational representation in the general population in the 1930s and 1990s. These ratios provide an even better indication of stability and change in denomination-

al representation in *Who's Who*. Ratios were calculated for all *Who's Who* listees and for those in the power elite and cultural elite. The 1926 census of religious bodies was used to determine denominational membership for the U.S. population during the earlier period (U.S. Bureau of the Census 1930). Since the census did not report Jewish congregational membership, Ament's (1927) estimate of Jewish membership was substituted.[5] We also used Ament's figures for Mormons in the 1926-27 *Who's Who* because Fry did not provide data on the number of Mormons in the 1930-31 edition. The combined 1988-91 General Social Surveys were used to determine denominational preferences for the general population in the 1990s (Davis & Smith 1991).

Findings

Overall, the results are quite consistent with our expectations. Though Episcopalians, Presbyterians, and UCCs as a group have lost some ground since the 1930s, they remain overrepresented among the nation's elite. They are slightly more prominent among power elites than cultural elites, though they are overrepresented in both spheres. Among other historical elites, Unitarian-Universalists are less prominent in the 1990s than they were in the 1930s but remain overrepresented in all spheres. Quakers have lost status among power elites but have gained among cultural elites. Catholics and Jews have made the most progress. Jews are now overrepresented among both power and cultural elites. In spite of Catholics' impressive gains, they remain underrepresented in both spheres. Among other Protestant groups, Baptists are most underrepresented of all and have declined since the 1930s. Disciples of Christ have also lost ground. Lutherans have gained, while United Methodists and the Reformed Church have had mixed experiences. Mormons have gained, and Christian Scientists have about the same stature as they did earlier.

Table 1 shows the religious affiliations of individuals in *Who's Who* in 1930-31 and 1992-93. In 1930-31, the Protestant Establishment accounted for 53.5% of all elites who reported a religious preference; in 1992-93, it accounted for 35.1% (a 34.4% decline). Episcopalians were the most tenacious of the three Protestant Establishment denominations; UCCs declined the most during the period. Episcopalians comprised 21.94% of the entries in 1930 and 18.04% in 1992 (not significant at .001). Presbyterians declined from 20.31% to 13.91% (a significant difference of 31.5%). UCCs declined from 11.29% to 3.19% (a 71.7% difference, which was significant).

Among other elite groups, Unitarians were significantly less prominent in the 1990s than they were in the 1930s. Although there were fewer Quakers among the elite, this decline was not significant at the .001 level.

Jews and Catholics gained the most (difference of proportion measures were statistically significant in both cases). Together, Jews and Catholics were only 5.76% of all *Who's Who* listees in 1930; by 1992, they were 35.44% (6.2 times as many). Jews comprised 1.31% of the entries in 1930, but 12.32% in 1992 (9.4 times as many).[6] Catholics rose from 4.45% of all entries in 1930 to 23.12% in 1992 (5.2 times as many).

TABLE 1: Reported Religious Affiliation of Individuals in *Who's Who*[a]

	1930-31	1992-93
Protestant establishment		
Episcopalian	21.94	18.04
Presbyterian	20.31	13.91*
Congregationalist/UCC	11.29	3.19*
Other elite		
Unitarian-Universalist	5.98	2.39*
Quaker	1.09	.65
Catholic	4.45	23.12*
Jew	1.31	12.32*
Other Protestant		
Baptist	8.97	4.71*
Disciples	2.05	.36*
Lutheran	2.41	6.01*
Methodist	14.50	9.57*
Reformed	1.00	.36
Other		
Mormon	.39[b]	1.52*
Christian Science	.69	.22
All others	3.62	3.63
	100.00	100.00
No affiliation listed	43.86	65.66

[a] The 1930-31 figures are reported by Fry (1933a).
[b] Reported by Ament (1927)

* 1930-92 difference significant $p < .001$

Among other religious groups, only Lutherans and Mormons increased in prominence. Baptists, Disciples of Christ, United Methodists, the Reformed Church, and Christian Scientists declined.

Table 2 shows the ratio of each denomination's representation in *Who's Who* to its representation in the U.S. church population (exact proportional representation is defined as 1.00). The Protestant Establishment groups were overrepresented in both the 1930s and the 1990s. Episcopalians are even more highly overrepresented among elites today than they were in the 1930s. In 1930 Episcopalians had 6.34 times as many listees among America's elite as we might

expect based on their distribution in the U.S. population. In 1992, they had 7.04 times as many listees. This increase was not matched either by Presbyterians, who declined from 3.35 to 2.75, or by UCCs, who declined even more (from 5.65 to 2.62). Despite these declines, Presbyterians and UCCs remain overrepresented among eminent Americans.

The other elite groups, the Unitarian-Universalists and Quakers, are still overrepresented. Unitarian-Universalists fell from a proportional ratio of 20.62 among the elites in 1930 to one of 9.60 in 1992. Quakers, on the other hand, are now more overrepresented than they were 60 years ago (up from 4.78 to 6.50).

The most dramatic increase was experienced by Jews. Though underrepresented in 1930 (only .73 as many entries as one might expect), Jews are now six times as likely to be represented among elites as they are to be found in the total population. Catholics also have gained. However, Catholics, who had only .13 times as many entries as one might have expected in 1930, were still underrepresented in 1992 (their ratio was .85).[7] Although Catholics have achieved parity with Protestants in overall socioeconomic status, they have not yet reached parity among elites.

With the exception of Christian Scientists (who are slightly overrepresented among elites and have about the same stature today as they did earlier), all other groups are underrepresented, as they were in the 1930s. There are proportionately more Lutherans, United Methodists, members of the Reformed Church, and Mormons among elites today than there were about sixty years ago. Baptists and Disciples, however, are less prominent now than they were in the 1930s.

Table 3 also shows denominational representation in *Who's Who* as a ratio proportional to general church membership, while distinguishing between power elites and cultural elites. Establishment groups remain overrepresented in both spheres and, as expected, are slightly more prominent among power elites than among cultural elites.

Episcopalians set themselves apart from other Establishment groups once again. Episcopalians were more prominent among power elites than among cultural elites in both periods. In 1930 they had 7.87 times as many leaders in business and public policy areas as might be expected based on their proportion of the total U.S. church membership; in 1992, they still had almost as many (7.41). There were 5.59 times as many Episcopalians as one might expect among leaders in cultural fields in the 1930s; in 1992, there were even more (6.44). Presbyterians also were more prominent among power elites than among cultural elites in both the 1930s and the 1990s. They had a proportional ratio of 3.60 among power elite entries in 1930; by the 1990s, they had a somewhat lower one (2.82). They had a proportional ratio of 3.23 among cultural elites in 1930; now they have one of 2.62. UCCs are different. In 1930, they were more overrepresented as cultural leaders (with a ratio of 6.20) than among power elites (with a ratio of 4.49). In 1992 they were about equally represented in both spheres (with ratios of 2.57 among business and political leaders and 2.53 among cultural leaders).

Unitarian-Universalists and Quakers were the most highly overrepresented of all religious groups among cultural elites. In 1992 these groups, respectively, had 14.8 and 12.7 times as many members among cultural elites as they had in

TABLE 2: Denominational Representation in *Who's Who* in Proportion to U.S. Church Membership

	1930-31[a]	1992-93[b]
Protestant establishment		
Episcopalian	6.34	7.04
Presbyterian	3.35	2.75
Congregationalist/UCC	5.65	2.62
Other elite		
Unitarian-Universalist	20.62[c]	9.60
Quaker	4.78	6.50
Catholic	.13	.85[d]
Jew	.73	6.00
Other Protestant		
Baptist	.55	.22
Disciples	.67	.32
Lutheran	.33	.83
Methodist	.88	.95
Reformed	.72	.93
Other		
Mormon	.33[e]	.66
Christian Science	1.28	1.22
All others	.64	.19

[a] U.S. membership figures are based on the proportion of members over 13 years of age reported in the 1926 census of religious bodies.
[b] U.S. membership figures are based on the proportion of General Social Survey respondents indicating a religious preference during 1988-91 ($N = 5{,}447$).
[c] Weighted mean based on the number of Unitarians and Universalists.
[d] Ament's (1927) estimate of Jewish congregational membership is used.
[e] Reported by Ament (1927).

the U.S. population as a whole. Though they were somewhat less prominent in the 1990s than in the 1930s, they remained the most culturally eminent groups in our society. Both groups also were overrepresented among power elites, but their prominence in economic and political arenas has waned considerably over the years.

Jews were slightly more prominent (though underrepresented) among power elites than cultural elites in the 1930s (.88 vs. .66). By 1992 Jews were noticeably more prominent in cultural fields than in business and politics (7.23

TABLE 3: Denominational Representation in *Who's Who* in Proportion to U.S. Church Membership

	1930-31	1992-93
	Power Elite	
Protestant establishment		
Episcopalian	7.87	7.41
Presbyterian	3.60	2.82
Congregationalist/UCC	4.49	2.57
Other elite		
Unitarian-Universalist	18.24	5.11
Quaker	3.87	1.30
Catholic	.15	.92
Jew	.88	5.17
Other Protestant		
Baptist	.46	.22
Disciples	.69	.34
Lutheran	.24	.79
Methodist	.82	1.04
Reformed	.48	.88
Other		
Mormon	.50[a]	.54
Christian Science	1.96	2.09
All others	.38	.15
	Cultural Elite	
Protestant establishment		
Episcopalian	5.59	6.44
Presbyterian	3.23	2.62
Congregationalist/UCC	6.20	2.53
Other elite		
Unitarian Universalist	21.62	14.81
Quaker	5.09	12.70
Catholic	.12	.74
Jew	.66	7.23
Other Protestant		
Baptist	.60	.22
Disciples	.65	.32
Lutheran	.37	.90
Methodist	.91	.82
Reformed	.81	.84
Other		
Mormon	.25[a]	.87
Christian Science	1.22	—
All others	.71	.23

[a] Reported by Ament (1927)

vs. 5.17). Catholics were about equally underrepresented in both spheres in the 1930s (.15 vs. .12). By the 1990s they were slightly more present among power elites than among cultural elites (.92 vs. .74).

Christian Scientists have been, and still are, more prominent among power elites than among cultural elites.[8] All other groups are underrepresented in both arenas, although Lutherans have gained considerable ground. Mormons also have gained. United Methodists and members of the Reformed Church have gained a bit among power elites, but their prominence among cultural elites has remained virtually the same. Baptists and Disciples of Christ have declined the most in both spheres.

Conclusions

Our data show that the Protestant Establishment has lost some ground relative to other groups since the 1930s, but contrary to the impression left by much of the literature declaring "the decline of the WASP," Establishment groups remain overrepresented by a considerable margin (by a factor of seven among Episcopalians and almost three among Presbyterians and UCCs). Episcopalians, Presbyterians, and UCCs no longer have the nearly exclusive access to positions of power, privilege, and prestige they once had. Their prominence in virtually all social hierarchies has declined a bit since the 1930s. However, Episcopalian influence has remained remarkably stable among power elites and has even increased among cultural elites. Though Presbyterians and UCCs have lost a bit more status, they remain overrepresented among the nation's power and cultural elites by two to three times their distribution in the total population.

The other elite groups, the Unitarian-Universalists and Quakers, are not as prominent as they used to be, but they remain overrepresented among both power and cultural elites (especially cultural elites, where they appear at about twice the Episcopalian rate).

The data also indicate that America's elites are more religiously diverse than they were about sixty years ago. While Episcopalians, Presbyterians, and UCCs were 53% of all elites in the early 1930s, they now comprise only 35%; other groups are now 65% of all elites. Jews and Catholics especially have made major strides in their climb toward the top of our society. Both groups have more access to power and privilege today than they ever had. Jews are now overrepresented among both power elites and cultural elites relative to their distribution in the total population. As we forecasted, they are more prominent among cultural elites than among power elites. Though Catholics have gained, they remain underrepresented in both spheres, especially among cultural elites.

Christian Scientists retain a presence among power elites. Lutherans have gained considerably in both spheres, as have Mormons, though both groups are underrepresented. United Methodists and members of the Reformed church have gained among power elites, but their prominence among cultural elites has remained stable. Baptists and Disciples of Christ have lost the most ground in both areas.

IMPLICATIONS

These findings have several implications. First, they are consistent with claims of writers such as Davidson (1994), Feagin and Feagin (1993), Greeley (1977), and Rossides (1990), who have emphasized the tenacity of WASP hegemony. Together, these studies suggest that members of the Protestant Establishment, and other religious elites, like most superordinate groups, want to maintain their power and privilege. Our results emphasize the continuing influence of groups that have played prominent roles in the shaping of the American way of life — a fact that seems to get lost among the louder voices declaring that Protestant hegemony has ended.

It is too early to declare, as writers such as Christopher (1989) have, that the Protestant Establishment is a thing of the past and that we are now living in a post-Protestant era. All three of the historically important Establishment groups remain overrepresented among both power and cultural elites. Even if their current rates of decline continue, these groups will have more than their fair share of influence in American society for several more decades.

Our study's emphasis on the different experiences of religious groups causes us to question the now-common claim that there is an ongoing convergence among Protestant denominations (Hammond 1992; Wuthnow 1988) and even among mainline religions (Roof & McKinney 1987). There is little evidence in this study to indicate a dramatic equalization of denominational status, as is claimed by convergence analysts. On the contrary, there is considerable evidence that denominational and religious family boundaries continue to be important factors that influence people's life chances.

These observations suggest two lines of further inquiry. There is a need for studies focusing attention on the way members of established religious groups maintain their prominence, especially among power elites. Studies of the methods by which members of the Protestant Establishment protect their advantages in the face of opposition would advance sociologists' understanding of how religion contributes to social inequality. To what extent do groups rely on mechanisms such as inheritance, legacy admissions, and appointments to high office to maintain their status? Do Establishment groups rely on different methods in business and politics than in cultural arenas such as higher education and the arts? We also need longitudinal studies of the implications of being among the elite. Do members of the Protestant Establishment who enjoy elite status today have the same opportunities to advance their self-interests that older generations of WASPs did? Or, as some writers claim, are they more limited in what they can do to maximize their own well-being and to impose their worldviews on others? Studies of the freedoms and constraints associated with elite status also would further our understanding of religion and social inequality.

Within an overall context of persistence, it also is true — probably as a result of the many developments we discussed in our theoretical section — that the Protestant Establishment no longer has the nearly exclusive control it once had over the nation's cherished resources. In addition to declining in the total population (cf. Finke & Stark 1992; Roof & McKinney 1987), it also is declining among the nation's power and cultural elite.

But the three Establishment groups, as well as the Unitarian-Universalists and Quakers, have had different experiences, which need to be examined more carefully in future studies. We need more research on how Episcopalians have held on most tenaciously, even gaining prominence in some areas, in the face of minority efforts to challenge their prominence and scholarly claims that denominations have lost their meaning. We also need additional studies of Presbyterians and UCCs to see how, despite their small numbers, they have preserved their influence, but also why they have failed to keep up with Episcopalians. How have Quakers gained in cultural prominence since the 1930s, while losing some ground in economic and political arenas?

Our findings also reaffirm many other studies indicating that Jews and Catholics have made important strides toward the top of our society. These two groups have come a long way since the 1930s. But perhaps more important, our study indicates how different these two groups' experiences have been. Jews, only 2% to 3% of the population, have made the most progress, going from marginal status among both power and cultural elites to being overrepresented in both areas, especially the cultural area. In the face of historical and lingering anti-Semitism, Jews have made remarkable gains and achieved eminence in both power and cultural circles. While Catholics, about one quarter of the U.S. population, have gained in overall socioeconomic status and in access to elite status, they still are underrepresented among elites. They may still struggle against subtle forms of anti-Catholicism, but at their current rate, they are likely to achieve parity in another decade or two. Comparative studies of these two groups — the opposition they have experienced and the gains they have made — would enhance our understanding of dominant-minority relations in general and religious superordination and subordination in particular.

Our results also suggest the need to study the other groups we have examined. Lutherans, who have gained the most among Protestant groups, and Baptists and Disciples, who have lost the most, would be of particular interest. How do Lutherans' success experiences compare with those of Catholics and Jews? What factors have contributed most to the decline of Baptists and Disciples?

Finally, this study documents important comparisons between the 1930s and 1990s. However, it does not tell us much about the rates of change during the intervening decades. Has the Protestant Establishment been able to maintain its prominence at approximately the same rate throughout all of the last five or six decades? How prominent were the Establishment groups in the 1950s compared to the 1970s? Is it true, as Christopher (1989) and others have suggested, that the Establishment lost ground most rapidly during the last 20 years? What about Jews and Catholics during the intervening years? Has their progress been steady, or uneven? Did they gain more rapidly between the 1930s and the 1950s, the 1950s and the 1970s, or the 1970s and the 1990s? How similar or different have their trajectories been during these decades? These questions need to be explored if we are to fully understand persistence and change in the Protestant Establishment and among other religious groups that will shape the American way of life in the years ahead.

Notes

1. Our analysis focuses on the three major groups composing the Protestant Establishment (Episcopalians, Presbyterians, and UCCs). However, we also will report data on two other religious groups that played major roles in shaping the nation's cultural orientations during the eighteenth and nineteenth centuries but have never been considered part of the Protestant Establishment: Unitarians (later Unitarian-Universalists) and Quakers (the Society of Friends).

2. Contrary to the emphasis of a status attainment explanation for mobility (Blau & Duncan 1967; Knottnerus 1987), a status persistence approach stresses the extent to which life chances are passed on across generations (Rossides 1990; Rytina 1989).

3. Persons selected on the basis of positions they hold include all members of Congress and the cabinet, federal judges, department heads, ambassadors, heads of the major universities, principal officers of the largest businesses, chief ecclesiastics of the major religious denominations, and others chosen because of incumbency in positions of responsibility (Marquis Who's Who 1992).

4. *Who's Who* entries were placed in the denominational categories shown in Table 1. The "all others" category includes groups with only a few listees and persons who indicated their affiliation as "Christian" or "Protestant."

5. The 1926 census figure for Jews is an estimate of the number of those of the Jewish faith living in communities where there was a congregation; consequently, this figure does not indicate the congregational membership (Engelman 1947; Goldstein 1981). Ament estimated that in 1927 Jews represented 1.8% of the total U.S. church membership.

6. Fry (1933b) argues that some Jews were probably reluctant to identify their religious preference in the 1930s. Assuming this is true, the real Jewish figure in 1930 might have been slightly higher than 1.31, and the actual difference between the 1930 and 1992 figures might be slightly smaller than the one reported in Table 1.

7. The Catholic index for 1930 is based on census figures showing that, after Jewish congregational membership is included, Catholics comprised about 34.2% of the U.S. church population. Ament's (1927) data suggest that this figure is inordinately high. We are inclined to think that Ament's estimate of Catholic membership (20.8% of the U.S. total in the mid 1920s) is a better approximation of the true Catholic figure. If Ament's estimate is substituted, the 1930 Catholic index of representation is .21 instead of .13.

8. The small number of Christian Scientists in the *Who's Who* sample accounts for their absence among cultural elites in 1992.

References

Alba, Richard D. 1990. Ethnic Identity. Yale University Press.

Alba, Richard D., and Gwen Moore. 1982. "Ethnicity in the American Elite." *American Sociological Review* 47:373-83.

Ament, William S. 1927. "Religion, Education, and Distinction." *School and Society* 26:399-406.

Anderson, Charles H. 1970. *White Protestant Americans.* Prentice-Hall.

Armstrong, Christopher F. 1990. "On the Making of Good Men: Character-Building in the New England Boarding Schools." Pp. 3-24 in *The High Status Track,* edited by Paul William Kingston and Lionel S. Lewis. SUNY Press.

Baltzell, E. Digby. 1958. *Philadelphia Gentlemen.* Prentice-Hall.

____. 1964. *The Protestant Establishment.* Random House.

____. 1976. "The Protestant Establishment Revisited." *American Scholar* 45:499-518.

_____. 1986. Foreword. Pp. ix-xiii in *Philadelphia Jewish Life, 1940-1985,* edited by Murray Friedman. Seth Press.

_____. 1991. *The Protestant Establishment Revisited.* Edited by Howard G. Schneiderman. Transaction.

Beale, Howard K. 1938. "Freedom for the School Teacher." *Annals of the American Academy of Political and Social Science* 200:119-43.

Blau, Peter M., and Otis D. Duncan. 1967. *The American Occupational Structure.* Wiley.

Carroll, Jackson, and Wade Clark Roof. 1993. *Beyond Establishment.* Westminster/John Knox.

Christopher, Robert. 1989. *Crashing the Gates: The De-WASPing of America's Power Elite.* Simon & Schuster.

Clark, Dennis. 1973. *The Irish in Philadelphia.* Temple University Press.

Cookson, Peter W., Jr., and Caroline Hodges Persell. 1985. *Preparing for Power: America's Elite Boarding Schools.* Basic.

Curtis, Lewis P., Jr. 1971. *Apes and Angels: The Irish in Victorian Caricature.* Smithsonian Institution Press.

D'Antonio, William V., James D. Davidson, Dean R. Hoge, and Ruth Wallace. 1989. *American Catholic Laity in a Changing Church.* Sheed & Ward.

Davidson, James D. 1985. "Theories and Measures of Poverty: Toward a Holistic Approach." *Sociological Focus* 10:177-98.

_____. 1994. "Religion among America's Elite: Persistence and Change in the Protestant Establishment." *Sociology of Religion* 55:419-40.

Davis, Beverly. 1953. "Eminence and Level of Social Origin." *American Journal of Sociology* 59:11-18.

Davis, James Allan, and Tom W. Smith. 1991. *General Social Surveys:1972-1991* [computer file]. Chicago: National Opinion Research Center [producer]. Storrs, Conn.: Roper Center for Public Opinion Research, University of Connecticut [distributor].

Domhoff, G. William. 1967. *Who Rules America?* Prentice-Hall.

_____. 1970. *The Higher Circles.* Random House.

_____. 1983. *Who Rules America Now? A View for the '80s.* Prentice-Hall.

Dye, Thomas R. 1983. *Who's Running America? The Reagan Years.* Prentice-Hall.

Engelman, Uriah Zvi. 1947. "Jewish Statistics in the U.S. Census of Religious Bodies (1850-1936)." *Jewish Social Studies* 9:127-74.

Feagin, Joe R., and Clairece Booher Feagin. 1993. *Racial and Ethnic Relations.* Prentice-Hall.

Finke, Roger, and Rodney Stark. 1992. *The Churching of America, 1776-1990: Winners and Losers in Our Religious Economy.* Rutgers University Press.

Fry, Luther C. 1933a. "The Reported Religious Affiliations of the Various Classes of Leaders Listed in *Who's Who,* 1930-1931 Edition." Pp. 311-14 in *Yearbook of American Churches, 1933.* Round Table Press.

_____. 1933b. "The Religious Affiliations of American Leaders." *Scientific Monthly* 36:241-49.

Gallup, George, and Jim Castelli. 1987. *The American Catholic People.* Doubleday.

Gaustad, Edwin S. 1968. "America's Institutions of Faith: A Statistical Postscript." Pp. 113-33 in *Religion in America,* edited by William G. McLoughlin and Robert N. Bellah. Beacon Press.

Glenn, Norval D. 1987. "The Trend in 'No Religion' Respondents to U.S. National Surveys, Late 1950s to Early 1980s." *Public Opinion Quarterly* 51:293-314.

Glenn, Norval D., and Ruth Hyland. 1967. "Religious Preference and Worldly Success: Some Evidence from National Surveys." *American Sociological Review* 32:73-85.

Goldstein, Sidney. 1981. "Jews in the United States: Perspectives from Demography." *American Jewish Year Book* 81:3-59.

Gordon, Milton M. 1964. *Assimilation in American Life.* Oxford University Press.

Greeley, Andrew. 1977. *The American Catholic*. Basic Books.

Hammond, Phillip E. 1992. *Religion and Personal Autonomy*. University of South Carolina Press.

Herberg, Will. 1960. *Protestant-Catholic-Jew: An Essay in American Religious Sociology*. Rev. ed. Doubleday.

Huntington, Ellsworth, and Leon F. Whitney. 1927. "Religion and 'Who's Who.'" *American Mercury* 11:438-43.

Hurst, Charles E. 1992. *Social Inequality*. Allyn & Bacon.

Hutchison, William R. 1989. "Protestantism as Establishment." Pp. 3-18 in *Between the Times*, edited by William R. Hutchison. Cambridge, England: Cambridge University Press.

Jaher, Frederic Cople. 1980. *The Urban Establishment*. University of Illinois Press.

Johnson, Benton. 1985. "Liberal Protestantism: End of the Road?" *Annals of the American Academy of Political and Social Science* 480:39-52.

Kalmijn, Matthijs. 1991. "Shifting Boundaries: Trends in Religious and Educational Homogamy." *American Sociological Review* 56:786-800.

Karabel, Jerome. 1984. "Status Group Struggle, Organizational Interests, and the Limits of Institutional Autonomy: The Transformation of Harvard, Yale, and Princeton 1918-1940." *Theory and Society* 13:1-40.

Karen, David. 1991. "'Achievement' and 'Ascription' in Admission to an Elite College: A Political-Organizational Analysis." *Sociological Forum* 6:349-80.

Keller, Suzanne. 1991. *Beyond the Ruling Class*. Transaction.

Kingston, Paul William, and Lionel S. Lewis. 1990. "Undergraduates at Elite Institutions: The Best, The Brightest, and the Richest." Pp. 105-20 in *The High-Status Track*, edited by Paul William Kingston and Lionel S. Lewis. SUNY Press.

Kirk, Dudley. 1957. "The Fertility of a Gifted Group: A Study of the Number of Children of Men in *Who's Who*." Pp. 78-98 in *The Nature and the Transmission of the Genetic and Cultural Characteristics of Human Populations*. Milbank Memorial Fund.

Kluegel, James R., and Eliot R. Smith. 1986. *Beliefs about Inequality*. Aldine de Gruyter.

Knottnerus, J. David. 1987. "Status Attainment Research and Its Image of Society." *American Sociological Review* 52:113-21.

Konolige, Kit, and Frederica Konolige. 1978. *The Power of Their Glory*. Wyden.

Larson, Cedric A. 1958. *Who: Sixty Years of American Eminence*. McDowell, Obolensky.

Levine, Steven B. 1980. "The Rise of American Boarding Schools and the Development of a National Upper Class." *Social Problems* 28:63-94.

Lichliter, Mary. 1946. "Social Obligations and Restrictions on Teachers." *School Review* 54:14-23.

Lieberson, Stanley, and Donna K. Carter. 1979. "Making It in America: Differences between Eminent Blacks and White Ethnic Groups." *American Sociological Review* 44:347-66.

Marger, Martin N. 1991. *Race and Ethnic Relations: American and Global Perspectives*. 2d ed. Wadsworth.

Marquis Who's Who. 1992. *Who's Who in America*. 47th ed. Author.

Mintz, Beth. 1975. "The President's Cabinet, 1897-1972: A Contribution to the Power Structure Debate." *Insurgent Sociologist* 5:131-48.

Moberg, David O. 1962. *The Church as a Social Institution*. Prentice-Hall.

O'Brien, Michael J. 1919. *A Hidden Phase of American History*. Devin-Adair.

Phillips, Kevin. 1990. *The Politics of Rich and Poor*. Random House.

Priest, T.B. 1982. "A Note on *Who's Who in America* as a Biographical Data Source in Studies of Elites." *Sociological Methods and Research* 2:81-88.

Pyle, Ralph E. 1995. "Persistence and Change in the Establishment: Religion, Education, and Gender among America's Elite." Ph.D. diss., Department of Sociology and Anthropology, Purdue University, West Lafayette, Indiana.

Reed, Julia. 1988. "The Changing of the Old Guard." *U.S. News and World Report* (8 Feb.):38-40; 45-47.

Roof, Wade Clark. 1979. "Socioeconomic Differentials among White Socioreligious Groups in the United States." *Social Forces* 58:280-89.

Roof, Wade Clark, and William McKinney. 1987. *American Mainline Religion*. Rutgers University Press.

Roozen, David A. 1993. "Denominations Grow as Individuals Join Congregations." Pp. 15-35 in *Church and Denominational Growth*, edited by David A. Roozen and C. Kirk Hadaway. Abingdon Press.

Rossides, Daniel W. 1990. *Social Stratification*. Prentice-Hall.

Ryan, William. 1981. *Equality*. Pantheon.

Rytina, Steve L. 1989. "Life Chances and the Continuity of Rank: An Alternative Interpretation of Mobility Magnitudes over the Life Cycle." *American Sociological Review* 54:910-28.

Schrag, Peter. 1970. *The Decline of the WASP*. Simon & Schuster.

Selth, Jefferson P. 1987. "National Prominence and Religious Preference." *Journal for the Scientific Study of Religion* 26:547-50.

Sklare, Marshall, and Joseph Greenblum. 1979. *Jewish Identity on the Suburban Frontier*. 2d ed. University of Chicago Press.

Sturdivant, Frederick D., and Roy D. Adler. 1976. "Executive Origins: Still a Gray Flannel World?" *Harvard Business Review* 54:125-32.

Sweet, William W. 1948. "The Protestant Churches." *Annals of the American Academy of Political and Social Science* 256:43-52.

Time. 1960. "Grosse Pointe's Gross Points." 75 (25 April):25.

Tropman, John E. 1986. "The 'Catholic Ethic' versus the 'Protestant Ethic': Catholic Social Service and the Welfare State." *Social Thought* 12:13-22.

U.S. Bureau of the Census. 1930. *Religious Bodies: 1926*. Vols. 1-2. Government Printing Office.

Verba, Sidney, and Gary R. Orren. 1985. *Equality in America*. Harvard University Press.

Weber, Herman C. 1933. "The 1932-1933 Edition of *Who's Who* and Organized Religion." Pp. 314-16 in *Yearbook of American Churches, 1933*. Round Table Press.

Weyl, Nathaniel. 1966. *The Creative Elite in America*. Public Affairs Press.

Wuthnow, Robert. 1988. *The Restructuring of American Religion: Society and Faith since World War II*. Princeton University Press.

Zweigenhaft, Richard L., and G. William Domhoff. 1982. *Jews in the Protestant Establishment*. Praeger.

____. 1991. *Blacks in the White Establishment?: A Study of Class and Race in America*. Yale University Press.

The Changing Basis of Competition in Organizational Populations: The Manhattan Hotel Industry, 1898-1990*

JOEL A.C. BAUM, *University of Toronto*

Abstract

Density-dependence theory cannot account for the widely observed pattern of proliferation and concentration in organizational populations. Although density dependence provides an explanation for initial proliferation, it cannot explain subsequent concentration because it does not allow some organizations to become dominant competitors. To address this basic limitation, I combine density dependence with three ecological models that permit size-based competitive asymmetries among a population's members and let the intensity of competitive processes vary over time. My analysis shows that Manhattan hotels of different sizes generate and experience different strengths of competition and that size-based competitive processes increase in strength over time, contributing to industry concentration. Separating empirically the effects of low-density conditions that occur early and late in a population's history also clarifies tests of density dependence in populations that have evolved beyond their peak density.

The evolutionary trajectories of diverse organizational populations appear to follow a common path. The number of organizations grows slowly initially, then increases rapidly to a peak. Once the maximum is reached, the number of organizations declines as one or a few large organizations come to dominate a much larger number of small organizations. Populations of newspaper publishers, labor unions, breweries, and banks, to list only a few, follow this basic pattern (Hannan & Carroll 1992). The application of ecological models of competition to explain this pattern has been a frequent topic in recent research in organizational sociology. An impressive array of evidence has been assem-

* *This research was conducted, in part, while the author was associate professor of management at the Stern School of Business, New York University. For comments on an earlier version of this article, I am grateful to Terry Amburgey, Jennifer Bethel, Martin Evans, John Freeman, Mike Hitt, Amy Hurley, Paul Ingram, Helaine Korn, Theresa Lant, Dan Levinthal, and anonymous* Social Forces *reviewers. This research benefited from my earlier collaboration with Steve Mezias on localized competition. For their data collection and coding efforts, thanks also go to Corrine Imbert, Bill Krause, Alan Mibab, and Sheila Peterson. Direct correspondence to the Division of Management and Economics, University of Toronto, Scarborough Campus, 1265 Military Trail, Scarborough, Ontario, M1C 1A4.*

bled linking the evolutionary trajectories of organizational populations to four ecological models of competition: *density dependence* (Hannan & Carroll 1992; Hannan & Freeman 1989), *mass dependence* (Barnett & Amburgey 1990), *size-localized competition* (Baum & Mezias 1992; Hannan, Ranger-Moore & Banaszak-Holl 1990), and *resource partitioning* (Carroll 1985; Carroll & Swaminathan 1992).

Among these models, density dependence is used most frequently in empirical research. This model focuses on population density, the number of organizations in a population, to explain the dynamics of mutualism (i.e., organizations enhancing each other's fates) and competition (i.e., organizations diminishing each other's fates) that underlie the growth trajectories of organizational populations. Although Hannan and Carroll (1992), Hannan and Freeman (1989), and others provide empirical support for the density-dependence model in a variety of organizational populations, density dependence provides an explanation only for the shape of the growth trajectory of organizational populations to their peak size, *not* for their subsequent domination by a few large organizations, i.e., concentration (Carroll & Hannan 1989). Density-dependence theory and research treat competition as a property of populations and assume that all of a population's members contribute to and experience competition equally (e.g., Winter 1990:286). This assumption ignores the asymmetries in competitive relationships that occur in most competitive situations (Carpenter et al. 1988). Since no organization or small group of organizations can dominate competition, the density-dependence model predicts logistic growth to an equilibrium level, but not the concentration that characterizes older populations. The tendency of organizational populations to become concentrated is of substantive interest, however, because it has implications for the structure of the distribution of power in organizational populations and communities (Hannan, Ranger-Moore & Banaszak-Holl 1990).

Notably, although density-dependence theory predicts logistic growth of populations only to their peak size, it is commonly tested in populations that have evolved well beyond their peak density (e.g., Hannan & Carroll 1992). This complicates density-dependence interpretations in two ways. One is that the meaning of population density varies greatly between early and later periods of a population's history. In early low-density conditions no organization or small group of organizations dominates and new entrants are typically similar to established organizations. By comparison, in late low-density conditions, one or a few large organizations often dominate substantial portions of the market, and organizations tend to differ considerably in size and strategy. Thus, increasing concentration diminishes the adequacy of population density as a proxy for the aggregate effects of a population on competition over the population's history (Winter 1990:289).[1] The other is that, when density is equally low early and late in a population's history, the legitimacy interpretation of the coefficient for the linear density term is unclear: Although early low density has a specific meaning in density-dependence theory (i.e., legitimation), late low density has no meaning at all.

The purpose of this article is to enrich our understanding of the influence of competitive processes on the evolutionary trajectories of organizational populations by combining density dependence with ideas from complementary models of ecological competition that allow different-sized organizations to

generate and experience competition of different strengths. Organizational size, arguably the dominant variable in organizational sociology (e.g., Scott 1987), is the most notable organizational characteristic affecting competitive processes neglected by density-dependent explanations for population dynamics. Mass dependence, size-localized, and resource partitioning models of ecological competition each address different aspects of the role of organizational size in competitive dynamics. Theory and research on these models suggests that the strength of competition generated and experienced by organizations in a population depends not only on the prevalence (i.e., density) of other organizations in the population, but on their sizes as well. By incorporating size-based competitive asymmetries, these models help explain the evolution of concentrated organizational size distributions.

In the original statement of density-dependence theory in organizational ecology, Brittain and Freeman (1980) identify important changes in the basis of competition that take place over the history of an organizational population. This article highlights the changing basis of competition, too, by examining how the strengths of competitive processes specified by density- and size-based models depend on the evolutionary point of the population. The predicted temporal variations in competitive processes contribute further to an explanation for the evolution of concentrated populations. I examine changes over time in the basis of competition in an analysis of organizational failure rates in the Manhattan hotel industry from 1898 to 1990.

Thus, in contrast to past ecological research, which treats the effects of competition as time-invariant (Isaac & Griffin 1989), I permit the effects of competitive processes to vary over a population's history. My approach also helps clarify tests of density dependence in populations that have evolved beyond their peak density by including variables that capture the effects of increasing organizational size variation on competitive processes directly in the analysis and by separating the effects of early and late low-density conditions empirically.

Ecological Models of Competition

DENSITY-DEPENDENT COMPETITION

The density-dependence model focuses on population density, the number of organizations in a population, to explain the dynamics of organizational populations. Hannan (1986) and his colleagues (Hannan & Carroll 1992; Hannan & Freeman 1989) propose that initial increases in population density produce mutualism by raising the institutional legitimacy of a population. The capacity of a population's members to acquire resources increases greatly when those who control resources take the organizational form for granted. However, as a population continues to grow, the interdependence among a population's members becomes competitive. When there are few organizations in a population, competition with others for scarce common resources can easily be avoided. But, as the number of potential competitors grows, avoidance becomes more difficult. Combined, the mutualistic effects of initial increases in density

and the competitive effects of further increases suggest curvilinear relationships between population density and organizational founding and failure rates.

Density-dependence theory predicts that growing density intensifies competition at an increasing rate. When there are few organizations in a population (relative to the abundance of resources), the addition of an organization increases the frequency and strength of competitive interactions only slightly, if at all, because competition with others for scarce common resources can easily be avoided. But as the number of population members grows (relative to the abundance of resources), avoidance becomes more difficult and the addition of an organization greatly increases the intensity of competition. Support has been found in a wide variety of empirical settings for the prediction that the intensity of competition increases at an increasing rate with population density (Hannan & Carroll 1992; Hannan & Freeman 1989; Singh & Lumsden 1990). Thus, the density-dependence model makes the following prediction:

Hypothesis 1a. Density dependence: The organizational failure rate increases at an increasing rate with increases in population density.

MASS-DEPENDENT COMPETITION

Density-dependence theory assumes that the intensity of competition experienced by an organization depends on the number of organizations in a population, regardless of their sizes. Small population members are assumed to place the same burden on the carrying capacity of the population as large members (Barnett & Amburgey 1990; Winter 1990). Yet there is broad agreement among scholars that large organizations secure a competitive advantage over smaller rivals as a result of their superior access to resources (e.g., Aldrich & Auster 1986; Starbuck 1965; Thompson 1967), greater market power (e.g., Bain 1956; Caves & Porter 1977; Edwards 1955; Pfeffer & Salancik 1978), and economies of scale and scope (e.g., Scherer & Ross 1990). If not all organizations in a population are equal competitors, then population density may not provide the most precise measure of competition faced by the organizations in a population (Baum & Mezias 1992).

To account for the greater potential competitive strength of a population's larger members, Barnett and Amburgey (1990) propose modeling the effects of population mass, the aggregate of the sizes of all organizations in a population (i.e., population density weighted by organizational size), along with population density. In contrast to density dependence, which examines whether larger *numbers of organizations* generate greater competition, mass dependence examines whether larger *organizations* generate greater competition. If larger organizations are stronger competitors, then, after accounting for the effects of population density, increases in population mass should have a competitive effect.

By permitting competitive strengths of organizations to vary as a function of size, mass-dependent competition allows larger organizations to dominate their populations. Thus, a key evolutionary implication of mass-dependent competition is that, over time, individual large organizations may displace, through competition, some of their population's size in numbers, resulting in

concentrated organizational size distributions (Barnett & Amburgey 1990:80). Large organizations may therefore be central to organizational ecology because they have a disproportionate influence on population dynamics (Barnett & Amburgey 1990).

Studies of mass-dependent competition present mixed findings. Some find competitive effects (Banaszak-Holl 1992; Baum & Mezias 1992; Hannan & Carroll 1992). Others find either no effect (Hannan & Carroll 1992) or a mutualistic effect (Barnett & Amburgey 1990). However, contradictory findings appear attributable to data limitations (Hannan & Carroll 1992:130-31) or significant features of the study population (Barnett & Amburgey 1990:98-99). The mass-dependent competition model makes the following prediction:

Hypotheses 2a. Mass dependence: The organizational failure rate increases with increases in population mass.

SIZE-LOCALIZED COMPETITION

Density-dependence theory assumes that each organization in a population competes with all other organizations in the population at an equal intensity (Baum & Mezias 1992). In the original statement of population ecology, however, Hannan and Freeman (1977:945-6) challenge this view. Following a long tradition in organizational sociology, they propose that organizations of different sizes use different strategies and structures. As a result, although organizations of different sizes are engaged in similar activities, large and small organizations depend on different mixes of resources. Therefore, Hannan and Freeman conclude that organizations compete most intensely with organizations of similar sizes. For example, if large and small organizations depend on different resources (e.g., large hotels depend on conventions while small hotels depend on individual travelers), then patterns of resource use will be specialized to segments of the size distribution. Consequently, competition between large and small organizations will be less intense than competition among large or among small organizations. Large organizations may, however, pose a threat to medium-sized organizations. Whatever strategy medium-sized organizations adopt to compete with large organizations will make them more vulnerable to competition from small organizations and vice versa. Therefore, the emergence of large organizations is likely to trigger a decline in the number of medium-sized organizations, while small ones flourish as their most intense competitors are removed from the environment.

Thus, analogous to the mass dependence model, a key evolutionary implication of size-localized competition is that, over time, individual large organizations may outcompete their smaller rivals, resulting in a more concentrated population. The main difference between the two models is that mass-dependent competition predicts the evolution of skewed (e.g., log-normal) size distributions (Barnett & Amburgey 1990) and size-localized competition predicts the evolution of bimodal size distributions (Hannan, Ranger-Moore & Banaszak-Holl 1990). Recent studies of Manhattan banks (Banaszak-Holl 1992), Manhattan hotels (Baum & Mezias 1992), and U.S. health maintenance organizations (Wholey, Christianson & Sanchez 1992) provide empirical evidence of size-

localized competition. The size-localized competition model makes the following prediction:

Hypothesis 3a. Size-localized competition: The failure rate of organizations of a given size increases with increases in the number of similarly sized organizations.

RESOURCE PARTITIONING

In work related closely to Hannan and Freeman's (1977) size-localized competition model, Carroll (1985) proposes that competition among large, generalist organizations in a population to occupy the center of the market frees peripheral resources that are most likely to be used by small, specialist members of the population. Carroll (1985) refers to the process generating this outcome as resource partitioning. The resource partitioning model implies that, in concentrated markets with a few large generalists, small specialists may be able to exploit more of the available resources without engaging in direct competition with larger, generalist organizations. This yields the prediction that increasing market concentration increases the failure rate of large, generalist organizations and lowers the failure rate of small, specialist organizations. Notably, the resource partitioning model offers one possible basis for size-localized competition and the evolution of concentrated, bimodal organizational size distributions: Large organizations secure the advantages of generalism, small organizations the advantages of specialism, and midsized organizations the liabilities of both (Meyer 1990). Thus, resource partitioning may constitute a subclass of size-localized competition. However, specification of concentration as a cause of size-localized competition provides a distinct insight.

In a study of newspaper organizations, Carroll (1985) found that, during the process of concentration, the failure rate of large, generalist newspapers increased while the failure rate of small, specialist newspaper organizations decreased, providing evidence that resource partitioning was at work. More recently, Carroll and Swaminathan (1992) found that as concentration increased in the American brewing industry, the failure rate of microbreweries decreased. However, the failure rate of large mass producers did not increase with market concentration. The resource partitioning model makes the following prediction:

Hypothesis 4a. Resource partitioning: As market concentration increases, the failure rate of large organizations increases and the failure rate of small organizations decreases.

Mass dependence, size-localized competition, and resource partitioning complement density-dependent explanations of population dynamics by highlighting two ways in which variation in organizational size creates competitive asymmetries that foster the evolution of concentrated size distributions common to older organizational populations. Mass dependence emphasizes that large organizations are dominant competitors and thus an especially important cause of competition. Size-localized competition and resource partitioning emphasize that size differences moderate interorganizational competition. Although the two roles large organizations play in aging populations appear contradictory, the prediction that larger organizations generate stronger competition is not incompatible with the idea that competition between large and small organizations is less intense than competition among large or

among small organizations. This is because increases in size-localized competition and resource partitioning are not incompatible with contemporaneous increases in the competitive strength of large organizations. Thus, size-localized competition and resource partitioning do not preclude the possibility that large organizations outcompete smaller rivals that operate in overlapping market segments.

Temporal Changes in the Basis of Competition

THE EVOLUTION OF ORGANIZATIONAL POPULATIONS: *r*-SELECTION, *K*-SELECTION, DIFFERENTIATION

The historical growth trajectories of diverse organizational populations appear to follow a common path: The number of organizations grows slowly initially, then increases rapidly to a peak. Once this peak is reached, there is a decline in the number of organizations and concentration increases (Hannan & Carroll 1992; Utterback & Suárez 1993). In their original statement of density-dependence theory, Brittain and Freeman (1980) identify several important changes in the basis of competition that accompany this common growth trajectory. According to Brittain and Freeman, organizational populations undergo a competitive transition from "*r*-selection" to "*K*-selection" as increasing population density alters the basis of competition from first-mover advantage to efficiency. They ground the distinction between these two bases of competition in the logistic growth functions underlying density-dependence theory in bioecology. A common logistic growth function is

$$\frac{dN}{dt} = rN[(K-N)/K]$$

where K is the carrying capacity of the population's environment, r is the natural growth rate of the population, N is the population density (i.e., the number of organizations in the population), and t is a time interval. In this equation, when population density is initially low, the natural growth term r dominates and organizations enter at a slow but exponentially increasing rate. However, as population density increases, the carrying capacity limit term K becomes dominant, and the rate of growth slows. Growth stops when population density is equal to the carrying capacity.

According to Brittain and Freeman (1980), in an emerging population, no organization or small group of organizations is dominant and new entrants are typically similar in size to established organizations. Competition is dominated by *r*-selection, which favors small organizations capable of moving quickly to exploit new resource opportunities in the resource-rich but dispersed and uncertain environments that characterize low-density conditions. Organizations compete by discovering new resource opportunities to exploit and by offering innovative goods and services. As the population continues to grow, its markets become connected, environmental uncertainty is reduced as demand becomes more predictable, and some organizations begin making broad, general appeals to all customers to expand their market shares. Along with the increased potential for competition that accompanies increasing population size (relative

to the abundance of resources), this begins to orient competition among the population's members toward cost and price reduction, favoring larger population members capable of achieving efficiency advantages.

When a population reaches the carrying capacity, the resources available to its members are exploited fully, and competitive pressures shift to *K*-selection, which favors larger organizations competing on the basis of efficiency. The economic and competitive advantages of the population's large members enable them to outcompete their smaller rivals, which are crowded out by the asymmetric competition. This "shakeout" decreases the number of organizations in the population, raises the level of market concentration, and creates a skewed or bimodal size distribution in which a few large organizations tend to dominate many smaller ones (Barnett & Amburgey 1990; Hannan & Freeman 1977; Ijiri & Simon 1977).

Ultimately, however, increasing market concentration forces large organizations to compete increasingly with one another for central markets capable of sustaining their large-scale operations (Carroll 1985). This specializes resource use to particular segments of the size distribution (Hannan & Freeman 1977) and creates pockets of demand that smaller, more specialized members of the population may be able to exploit without engaging in direct competition with the larger generalists (Barnett & Amburgey 1990; Carroll 1985). Thus, resource partitioning localizes the competitive effects of organizations — even large ones — to their particular segments of operations. Competition at carrying capacity also initiates a more general process of functional and territorial differentiation within the population as competition pushes entrepreneurs to seek out distinct functions in which they hold a competitive advantage (Hawley 1950). This creates organizational subgroupings that fulfill complementary roles in which they are dependent on, but noncompetitive with, each other (Baum & Singh 1994a, 1994b) and lowers the potential for competition by reducing the number of direct competitors each population member confronts (Baum & Mezias 1992; Delacroix, Swaminathan & Solt 1989).[2]

TEMPORAL HYPOTHESES

The foregoing account, which characterizes the evolution of organizational populations as a three-stage process of *r*-selection, *K*-selection, and differentiation, suggests that the strength of each of the four competition processes will change substantively over the course of an organizational population's history. First, if the shift to competition on the basis of efficiency (i.e., *K*-selection) increases the competitive strength of large organizations over their smaller rivals, then the strength of mass-dependent competition should increase over a population's history. Second, if the number of direct competitors (i.e., the number of other organizations that require the same resources) each organization faces within the population is reduced by resource partitioning and organizational differentiation, then density-dependent competition should weaken over a population's history. Third, if increased competition among large organizations for a population's central markets leaves smaller organizations to compete among themselves for more diverse, peripheral markets, then size-localized competition should intensify over a population's history. Fourth, and

finally, if large organizations focus increasingly on competing for central markets on the basis of efficiency, then the competitive effect of concentration on large organizations should increase over a population's history. These ideas suggest the four following hypotheses:

Hypothesis 1b. Density dependence: Density-dependent competition weakens over the course of a population's history.

Hypothesis 2b. Mass dependence: Mass-dependent competition strengthens over the course of a population's history.

Hypothesis 3b. Size-localized competition: Size-localized competition intensifies over the course of a population's history.

Hypothesis 4b. Resource partitioning: The competitive effect of increasing market concentration on large organizations increases over the course of a population's history.

Thus, while the influence of density-dependent competition is predicted to decline over a population's history, mass dependence, size-localized competition, and resource partitioning are predicted to play increasingly influential roles. If these size-based competitive processes are found to increase in strength over time, this would provide new insights into the propensity of older organizational populations to evolve concentrated size distributions.

Methods

DATA DESCRIPTION

The data used in this study include life history information on all 571 transient hotels that operated in Manhattan at any time between 1898 and 1990. Transient hotels are those catering to short-term visitors as opposed to residential hotels that serve long-term or permanent guests. Residential hotels were excluded from the sample. Four archival sources were used to construct these life histories. The *Hotel Red Book,* published annually since 1887, contains detailed information on the name, number of rooms, location, and room rates of hotels. It is the most comprehensive historical listing of Manhattan hotels in existence. The information contained in the *Red Book* was cross-referenced and supplemented using three additional archival sources: (1) the *Manhattan Classified Directory/Yellow Pages,* published since 1929; (2) the *Annual Directory of the Hotel Association of New York City,* published since 1940; and (3) the *Hotel and Travel Index,* published since 1951. Because detailed organizational data are missing for many hotels prior to 1898, the observation period for this study begins in 1898 even though the archival sources begin in 1887.

Between 1898 and 1990 the Manhattan hotel industry gradually transformed from one dominated by entrepreneurial, independent operators to one dominated by professionally managed hotels. After a period of slow growth in the decades preceding and following the turn of the century, the Manhattan hotel industry grew rapidly alongside the expansion of Manhattan during the "roaring twenties," reaching a maximum of 353 hotels in 1931. Since then, the industry has become more concentrated, a trend that accelerated after 1960 as smaller and medium-sized hotels were replaced by fewer and larger establish-

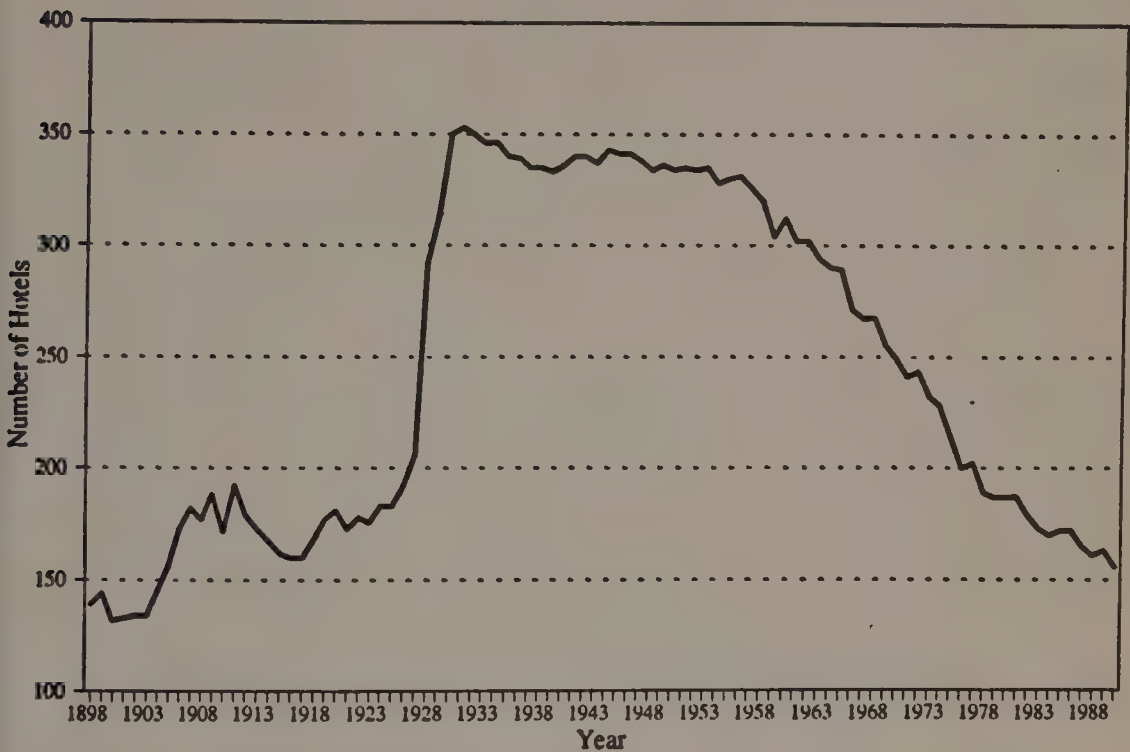
ments. The number of hotels in Manhattan in each year is shown in Figure 1.[3] Figure 2 illustrates how the average size of a Manhattan hotel, measured by the number of rooms, increased from fewer than 250 in 1898 to more than 400 in 1990.

During the study period, 438 transient hotels were founded in Manhattan. A hotel was defined to have been founded in the year it first appeared in any archival source. Furthermore, 339 (77.4%) of the hotels founded in Manhattan since 1898 had ceased operations by the end of 1990. Failure was defined as the cessation of hotel services for short-term visitors. Changes in name or ownership were not included as failures because the organization itself continued to provide transient-hotel services. The date of failure was defined as the year a hotel was permanently delisted from the *Red Book.* Whenever *Red Book* listings differed from those in the other archival sources, a hotel was defined to have failed in the first year it was permanently delisted from two or more of the archival sources.

In 1898, 133 of the hotels in the sample were already in operation; thus, the life histories for these hotels were left-censored (i.e., founded before the study period). To include as much of the data on hotels from the 1898-1990 period in the study as possible, I attempted to find founding dates for hotels existing in 1898. With available archival information, I was able to confirm founding dates for 112 (84.2%) left-censored hotels. Because their ages were not known, the 21 hotels with unknown founding dates could not be included in the analysis, although information on these hotels was included in the computation of population density, population mass, localized competition, and market concentration described below. Therefore, the final sample for the analysis included 550 hotels, of which 394 (71.6%) failed.

The competitive strengths of large hotels stem from several potential sources. Large hotels derive economies of scope from the ability to offer a wider range of services and to cater to a larger variety of clients (e.g., business travelers, corporate meetings, conferences, tourists, tour groups) simultaneously. This generalist capability provides shelter from environmental changes that affect demand in one client group but not in others. Large hotels can also exploit economies of scale in conference, dining, and recreational facilities that allow them to operate more efficiently and offer lower prices as a result of lower per-room costs than is possible for smaller hotels. Economies of scale are also realized in purchasing, reservation systems, advertising, marketing, and promotion. Large hotels are also better equipped to undertake major renovations by applying more resources to the task and maintaining the hotel's activities during periods of alteration by, for instance, closing several floors for renovation. Large hotels are also more "rate flexible," that is, capable of using predatory pricing to outcompete their smaller competitors for scarce clients. As the manager of one small Manhattan hotel put it, "If [Hotel X] put 100 of its 1,400 rooms on the market for $99, I'd be dead in the water. If I cut the rate on all my rooms, [Hotel X] wouldn't even notice, and I'd be out of business in no time." Rate flexibility also provides large hotels an advantage with corporate clients.

FIGURE 1: Number of Manhattan Hotels

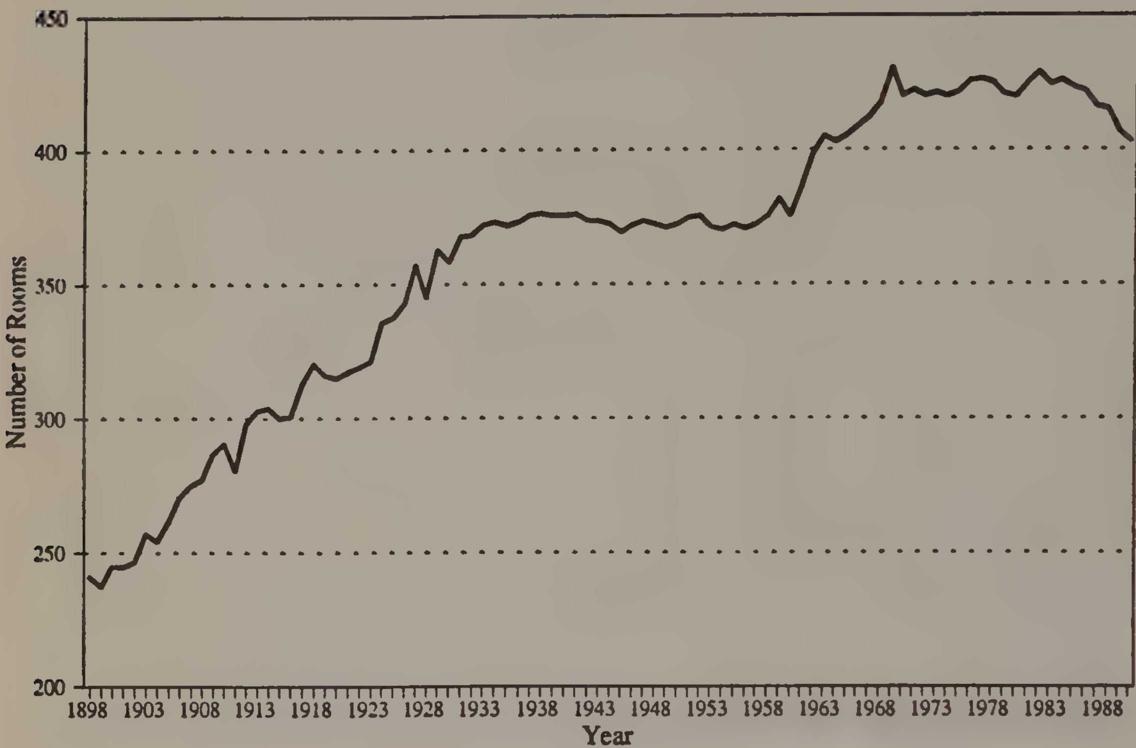


INDEPENDENT VARIABLES

Population density was measured as the total number of hotels existing at the start of each year. To allow for the curvilinear effect predicted by density-dependence theory, density was modeled as a quadratic function by including both density and density squared (divided by 1,000 for rescaling) in the analysis. Carroll and Hannan (1989) suggest that failure to incorporate data on the early period of a population's history may obscure the legitimating effects of population density. Consequently, because data on the Manhattan hotel industry are missing prior to 1898, the competitive effects of population density may be expected to dominate the density estimates. Density-dependent competition (hypothesis 1a) implies a positive effect of population density squared on the failure rate.

In order to capture the differential competitive strength of larger organizations, a measure of organizational size must capture total productive capacity, or plant size (Winter 1990). For hotels, the number of rooms operated represents this "productive capacity installed" (Winter 1990:288). Therefore, I derive all size-related measures based on the number of rooms a hotel operates at the start of each observation year, which is also the standard measure of size used in the hotel industry (Wyckoff & Sasser 1981).

FIGURE 2: Average Manhattan Hotel Size



Population mass was measured as the total number of rooms in all hotels (i.e., the total productive capacity of the industry) existing at the start of each year.[4] Following Barnett and Amburgey (1990), each hotel's number of rooms was subtracted from total mass, so the mass variable reflected the number of rooms operated by other hotels, and a natural logarithm transformation of mass was used. With density controlled, the effects of changes in the mass variable capture the effects of changes in the *average size* of other organizations (Barnett & Amburgey 1990:92). Therefore, mass-dependent competition (hypothesis 2a) requires a positive effect of population mass on the failure rate.

Size-localized competition was measured by comparing the size of a focal hotel, S_{it}, to the sizes of all other hotels in Manhattan, S_{jt}, $i \neq j$, within μ units of its own size at time t.[5] This is consistent with Hannan and Freeman's (1977) idea that competition is localized to particular segments of the organizational size distribution. The size differences between the focal hotel and the hotels within its competitive range were converted to a Euclidean distance, as follows:

$$D_{it}^{size \pm \mu} = \sqrt{\sum_{|S_{it}-S_{jt}|<\mu}^{\mu} (S_{it}-S_{jt})^2}. \tag{2}$$

$D_{it}^{size \pm \mu}$ increases with the average size difference between the focal hotel and other hotels within μ units, while the distances to hotels not within μ units of the focal hotel have no effect. Thus, large values for $D_{it}^{size \pm \mu}$ imply less potential size-localized competition. Therefore, size-localized competition (hypothesis 3a) presupposes a negative effect of $D_{it}^{size \pm \mu}$ on the failure rate.

The value for μ was set to the size of the focal hotel divided by two (i.e., $\mu = \pm S_i/2$) based on the earlier analysis of these data by Baum and Mezias (1992). This variable is called $D_{it}^{size/2}$ in the analysis.[6] This setting implies that a 100-room hotel competes with hotels ranging from 50 to 150 rooms, while a 1,000-room hotel competes with those from 500 to 1,500 rooms. Thus, $D_{it}^{size/2}$ limits the range of sizes over which hotels compete and takes into account that the competitive reach of large organizations is wider than that of small organizations (Winter 1990).

Last, *market concentration* was measured using the eight-firm concentration ratio, that is, the share of the total industry capacity accounted for by the eight largest hotels at the start of each year. The resource partitioning model predicts that increasing market concentration increases failure rates for large organizations but decreases failure rates for small organizations. Therefore, in addition to the main effect of market concentration, the effect of the interaction of market concentration and the size of a hotel was estimated. Support for resource partitioning (hypothesis 4a) requires a negative main effect for market concentration and a positive interaction between market concentration and size.

INDUSTRY AGE AND INTERACTION EFFECTS

Hypotheses 1b-4b predict variation in the effects of density, mass, size-localized competition, and market concentration over a population's history. To test these hypotheses, interaction terms were created to examine how the effects of the variables for the competitive processes changed over time. To do this, an *industry age* variable was constructed and interacted with the variables for the competitive processes. Industry age was set equal to the number of years since 1793, the year in which the City Hotel, the first modern hotel, was founded in Manhattan (Batterberry & Batterberry 1973).[7] Support for hypothesis 1b, which predicts a decline in the competitive effect of population density over time, requires a negative effect of the industry age × density squared interaction on the failure rate. Support for hypothesis 2b, which predicts an increase in the competitive effect of population mass increases over time, requires a positive effect of the industry age × mass interaction on the failure rate. Support for hypothesis 3b, which predicts an increase in size-localized competition over time, requires a negative effect of the industry age × $D_{it}^{size/2}$ interaction on the failure rate (recall that increases in $D_{it}^{size/2}$ indicate lower potential size-localized competition). Lastly, support for hypothesis 4b, which predicts an increase in the competitive effect of market concentration on large organizations over time, requires a positive effect of the industry age × size × market concentration interaction on the failure rate.

As is evident from Figure 1, the Manhattan hotel industry is well past its peak density. As described earlier, this complicates the legitimacy interpretation of the coefficient for the linear density term because late low density has no role

in density-dependence theory. This problem is corrected empirically by separating the effects of early low density from those of late low density. I do this by interacting the linear density term with a quadratic specification for the industry age variable, i.e., including separate terms for industry age and industry age squared (divided by 1,000 for rescaling). Because the squared term dominates the quadratic effect when industry age is high, the industry age interaction term captures the early low-density effect, and the industry age squared interaction captures the late low-density effect. Although density-dependence theory is commonly tested in populations that have evolved beyond their peak density (e.g., Hannan & Carroll 1992), effects of early and late low density are not separated in past research.[8]

CONTROL VARIABLES

To rule out plausible alternative explanations, I control for the effects of several organizational characteristics and variables influencing the environmental carrying capacity.

Organizational Variables

Age was defined as the number of years since the date of a hotel's founding. *Size* and *price* were measured, respectively, as the number of rooms operated by a hotel and the average daily room rate (in constant dollars) advertised by a hotel in each year of its existence. These variables were constructed using information contained in the archival sources described above. For a small number of hotels for which size or price data were unavailable for a given year, linear interpolation was used to determine the values. The logarithms of age, size, and price were used to reduce the variables' skewness. A left-censored dummy variable, coded 1 for hotels founded before 1898 and 0 otherwise, was also included to examine whether the hotels founded before 1898 had systematically different failure rates.[9]

Environmental Variables

The study also controlled for several factors influencing the environmental carrying capacity for transient-hotel services in Manhattan. The *potential demand* for hotel services was measured as the number of visitors (divided by 1,000,000 for rescaling) to New York City in the prior year. This variable includes arrivals by sea, rail, and air. Because the hotel industry is vulnerable to the state of the economy (Wyckoff & Sasser 1981), the annual *gross national product (GNP) growth rate* was included as a control.[10] The effects of the *number of foundings and failures* in the prior year on current failure rates were also controlled (Delacroix, Swaminathan & Solt 1989).

ANALYSIS

The failure rate of Manhattan hotels is estimated using $r(t)$, the instantaneous rate of failure. The hazard rate of a hotel failing is defined as:

$$r(t) = \lim_{\Delta t \to 0} [Pr(t, t+\Delta t|t)/\Delta t] \quad (3)$$

where $Pr(t, t+\Delta t \mid t)$ is the probability of failure in the interval t, $t+\Delta t$ given the hotel was still alive at time t. The failure rate was modeled using the following specification:

$$r(t) = \exp[\alpha X(\tau)],\ \tau_{(j-1)} < t \leq \tau_j \quad (4)$$

where $X(\tau)$ is a vector of covariate values at time τ, and α is a vector of coefficients. In this exponential model, transition rates are postulated to be log-linear functions of the variables in X. To include time variation in the covariates, a multiple-spells formulation was used. In this formulation of the model, each hotel's history is broken down into one-year spells in which the hotel is at risk of failing. Each spell is considered right-censored unless the hotel fails. This allows the time-varying independent variables to be updated annually. *RATE* (Tuma 1980) was used to estimate the vector of parameter estimates α by the method of maximum likelihood.

Results

Table 1 reports maximum-likelihood estimates for the analysis of Manhattan hotel failure rates. Model 1 provides a baseline model of competitive processes that includes a quadratic specification of industry age and the organizational and environmental control variables. Model 1 provides mixed support for the ecological models of competition. Coefficients for the quadratic specification of population density are both significant but have signs opposite in direction to those predicted by density-dependence theory. In contrast to hypothesis 1a, this implies that competition among hotels was weakest when population density was highest and strongest when population density was lowest. The mass dependence model (hypothesis 2a) is, however, supported by the significant, positive coefficient for population mass. With population density controlled, this means that increases in the average size of other hotels had competitive effects that increased the failure rate for smaller hotels. The coefficient for $D_{it}^{size/2}$ is significant and negative as predicted by the size-localized competition model (hypothesis 3a). This indicates that hotels with smaller average Euclidean distances from others within their competitive size range exhibited higher failure rates. In other words, hotels with more similar-sized competitors exhibited significantly higher failure rates. However, the resource partitioning model (hypothesis 4a), which predicts that the failure rate of large organizations increases with increases in market concentration, is not supported. Together, the significant, positive coefficient for market concentration and the significant, negative estimate for its interaction with size indicate that large size-buffered hotels from the hazardous effects of market concentration.

Model 2 introduces the interactions between industry age and population density. Adding these interaction terms improves the fit of the model significantly ($\Delta\chi^2 = 17.70$, df $= 3$, $p < .05$). In model 2, the coefficients for the quadratic population density specification are both significant and positive. The positive population density squared coefficient supports the density-dependent competi-

TABLE 1: Maximum Likelihood Models of Manhattan Hotel Failure, 1898-1990[a]

	Model				
	1	2	3	4	5
Constant	11.33	-27.99	-209.9*	-211.2*	-153.0
	(12.75)	(34.69)	(123.0)	(123.0)	(128.9)
Log (age)	.357*	.380*	.384*	.385*	.414*
	(.187)	(.187)	(.186)	(.187)	(.188)
Log (size)	.025	-.253	-.475	-.485	-2.522
	-(.422)	(.445)	(.467)	(.466)	(1.537)
Log (price)	1.864*	-1.815*	-1.825*	-1.828*	-1.789*
	(.259)	(.259)	(.259)	(.259)	(.258)
Left-censored	-.019	-.026	-.026	-.017	-.034
	(.224)	(.224)	(.224)	(.224)	(.224)
Potential demand	-.035*	-.038*	-.030*	-.030*	-.027*
	(.010)	(.013)	(.013)	(.013)	(.014)
GNP growth rate	-.035*	-.035*	-.024*	-.024*	-.024*
	(.010)	(.011)	(.011)	(.011)	(.011)
Number of prior foundings	.008	.000	.000	.000	.002
	(.006)	(.008)	(.008)	(.008)	(.008)
Number of prior failures	-.045*	-.048*	-.051*	-.051*	-.050*
	(.013)	(.013)	(.013)	(.013)	(.013)
Industry age	-.728*	.879*	1.844*	1.853*	1.144
	(.108)	(.459)	(.776)	(.777)	(.804)
Industry age^2/1,000	2.423*	-3.170*	-1.485	-1.486	-1.126
	(.334)	(1.505)	(1.854)	(1.854)	(1.864)
Population density	.043*	.457*	.535*	.533*	.513*
	(.026)	(.171)	(.282)	(.282)	(.282)
Population density^2/1,000	-.098*	.551*	.751*	.755*	.639*
	(.038)	(.156)	(.338)	(.338)	(.347)

[a] Standard errors in parentheses. The sample included 22,042 annual spells and 394 failures.

* $p < .05$

tion prediction (hypothesis 1a). In addition, the coefficient for the industry age × density squared interaction is significant and negative, supporting hypothesis 1b. Taken together, these estimates indicate that the failure rate increased at an increasing rate with population density but that the competitive effect of increasing population density weakened over the hotel industry's history. The estimates for the linear density term's interactions with quadratic specification for industry age are also noteworthy. The industry age × density interaction is

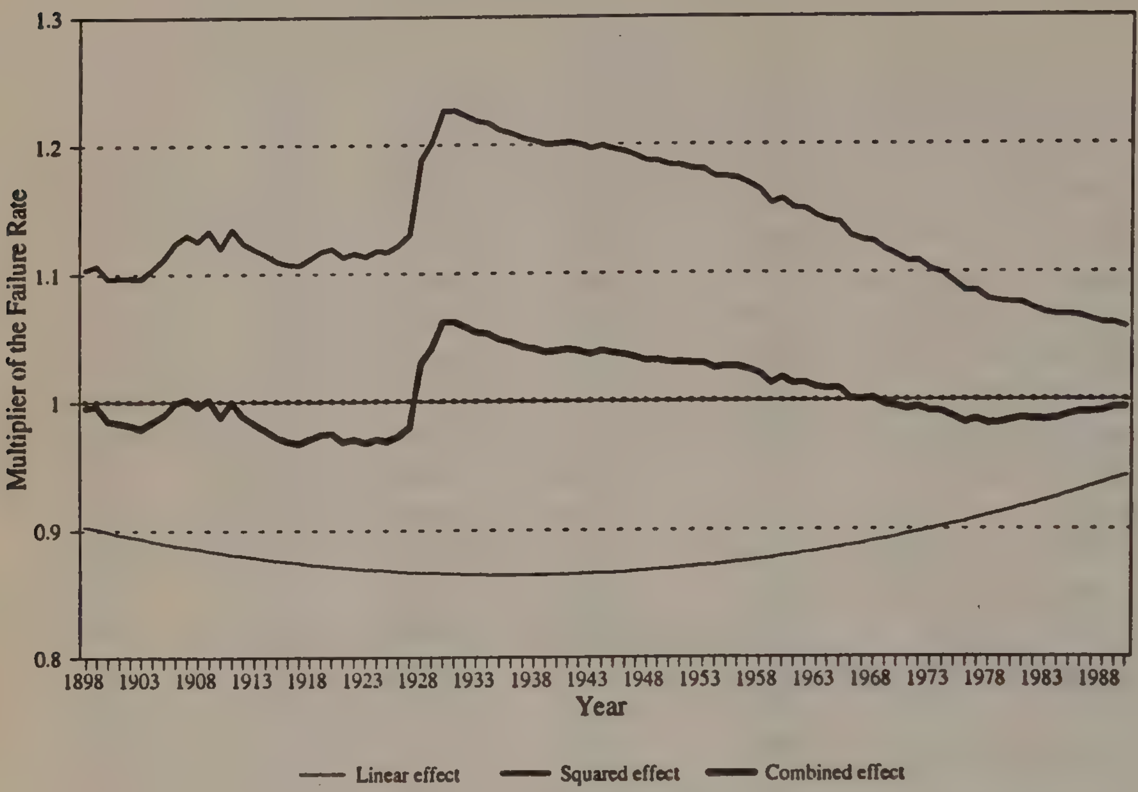
TABLE 1: Maximum Likelihood Models of Manhattan Hotel Failure, 1898-1990[a]

	Model				
	1	2	3	4	5
ln (Population mass)	3.150*	-3.873	-20.45*	-20.57*	-21.18*
	(1.656)	(2.256)	(6.341)	(6.431)	(6.611)
$D_{it}^{size/2}/100$	-.117*	-.116*	-.116*	-.065	-.063
	(.068)	(.068)	(.068)	(.079)	(.080)
Market concentration	18.28*	19.89*	8.499	8.383	25.42*
	(8.919)	(9.376)	(12.10)	(12.09)	(14.03)
Log (size) x Market concentration	-10.89*	-11.67*	-14.92*	-14.80*	-14.84*
	(4.601)	(4.913)	(5.212)	(5.208)	(6.543)
Industry age x Density		-.0084*	-.0074*	-.0074*	-.0072*
		(.0023)	(.0032)	(.0032)	(.0032)
Industry age^2/1,000 x Density		.0293*	.0258*	.0258*	.0249*
		(.0077)	(.0080)	(.0080)	(.0081)
Industry age x Density^2/1,000		-.0019*	-.0044*	-.0044*	-.0039*
		(.0010)	(.0025)	(.0025)	(.0026)
Industry age x ln (Mass)			.103*	.104*	.113*
			(.025)	(.025)	(.028)
Industry age x Density^2/100				-.0031*	-.0036*
				(.0019)	(.0019)
Industry age x log (Size)					.020
					(.017)
Industry age x Market concentration					-.604*
					(.151)
Industry age x log (Size) x Market concentration					.279*
					(.131)
χ^2	295.58	313.28	323.66	327.66	347.68
Degrees of freedom	16	19	20	21	24

significant and negative and the industry age squared × density interaction is significant and positive. Combined with the positive linear effect of population density, the interaction effects indicate that density increases lowered the failure rate in early low-density conditions but raised the failure rate in late low-density conditions.

Figure 3 illustrates the temporal variation in the population-density and population-density-squared effects graphically. This figure gives multipliers that show the effects of adding one hotel to population density in each year. In the figure, a multiplier of greater than one indicates that the failure rate increases relative to the estimate based on the actual density in that year by a factor equal

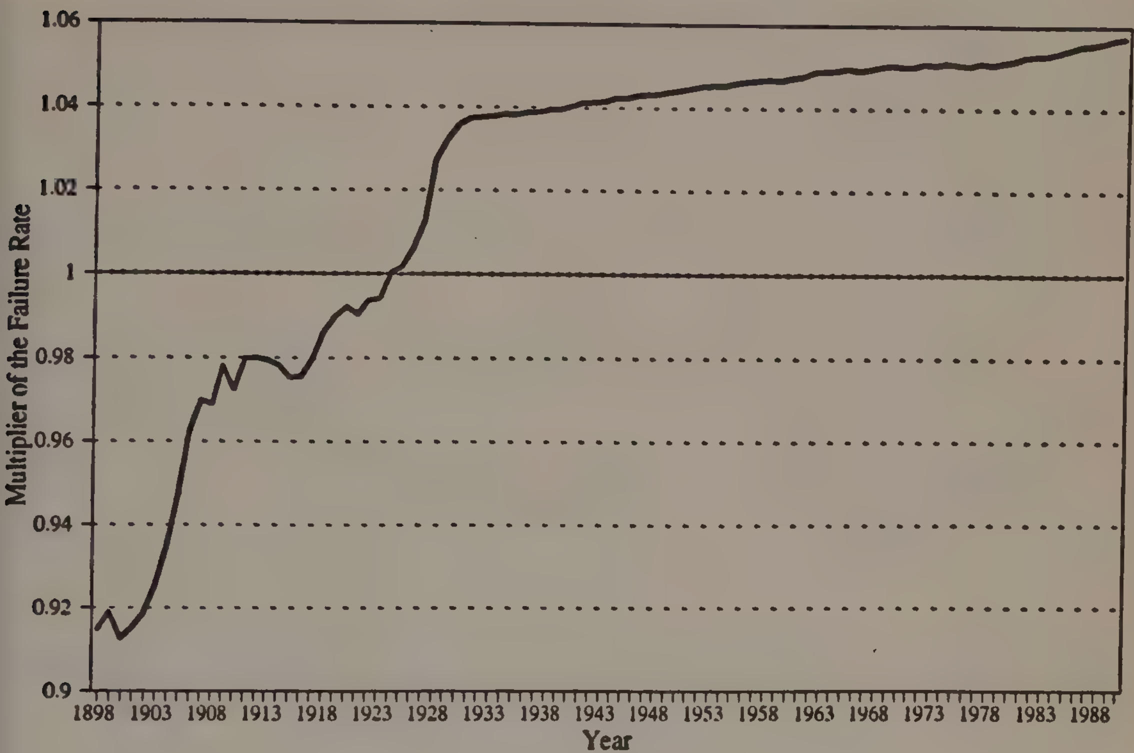
FIGURE 3: Population Density Effects (Model 2)



to the multiplier. A multiplier of less than one indicates that the failure rate decreases by such an amount. As the figure shows, the multiplier for the squared (i.e., competitive) effect of adding one hotel hovers around 1.1 from 1898 to 1927, jumps to over 1.12 in 1931 along with the industry's rapid 1928-31 expansion, and declines steadily thereafter as predicted by hypothesis 1b. By comparison, the multiplier for the linear effect of adding one hotel to the population declines initially from .9 to .86 but then increases to .94 by the end of the observation period. The relatively small initial decline may result, as noted above, from missing data on the population's early history, obscuring the legitimating effects of population density (Carroll & Hannan 1989). Nevertheless, consistent with the density-dependent legitimation argument (Hannan & Carroll 1992), the multiplier for the overall effect shows that density increases in early low-density conditions lowered failure rates in the Manhattan hotel industry. Thus, only after incorporating time variation in the effects of population density and population density squared are density-dependent processes of legitimation and competition observed clearly.

Model 3 introduces the industry time clock × mass interaction. Adding this interaction term again improves the fit of the model significantly ($\Delta\chi^2 = 10.47$, df = 1, $p < .05$). In model 3, supporting hypothesis 2b, the coefficient for the industry age × mass interaction is significant and positive. Moreover, in contrast to the estimates for population mass in models 1 and 2, the estimate for the

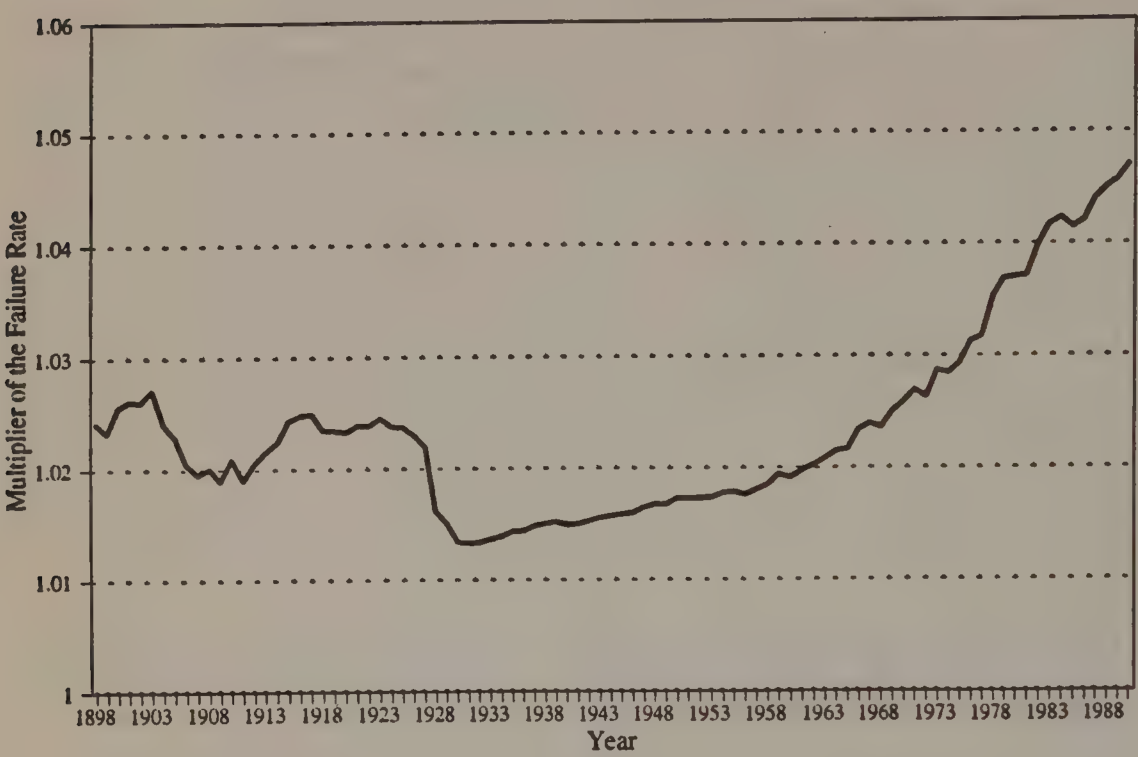
FIGURE 4: Population Mass Effects (Model 3)



main effect of mass is significant and negative in model 3. In combination, with population density controlled, these estimates for the main and time-varying effects of population mass indicate that, early in the observation period, increases in the average size of other hotels had mutualistic effects that lowered the failure rate. However, over the observation period, increases in the average size of other hotels had increasingly competitive effects that increased the failure rate of smaller hotels. Thus, unless time variation is considered, the mass estimates fail to distinguish the early mutualistic and later competitive effects of large hotels. This temporal variation in the mass effect is illustrated in Figure 4, which gives a multiplier that shows the effect of adding 500 rooms to population mass in each year. As Figure 4 shows, the multiplier for adding 500 rooms to the population increases rapidly from 0.92 in 1898 to 1.04 in 1930, increasing more moderately thereafter. Notably, the multiplier is less than 1 before 1925, indicating that increases in population mass had a mutualistic effect that lowered the failure rate early in the observation period. Thus, mass-dependent competition played its role in the evolution of the industry later in the observation period as large hotels gained competitive strength, displacing smaller hotels and decreasing the number of hotels.

Introducing the industry age × $D_{it}^{size/2}$ interaction in model 4 improves the fit of the model significantly again ($\Delta\chi^2 = 4.00$, df = 1, $p < .05$). In contrast to previous models, the estimate for $D_{it}^{size/2}$ is not significant in model 4. However,

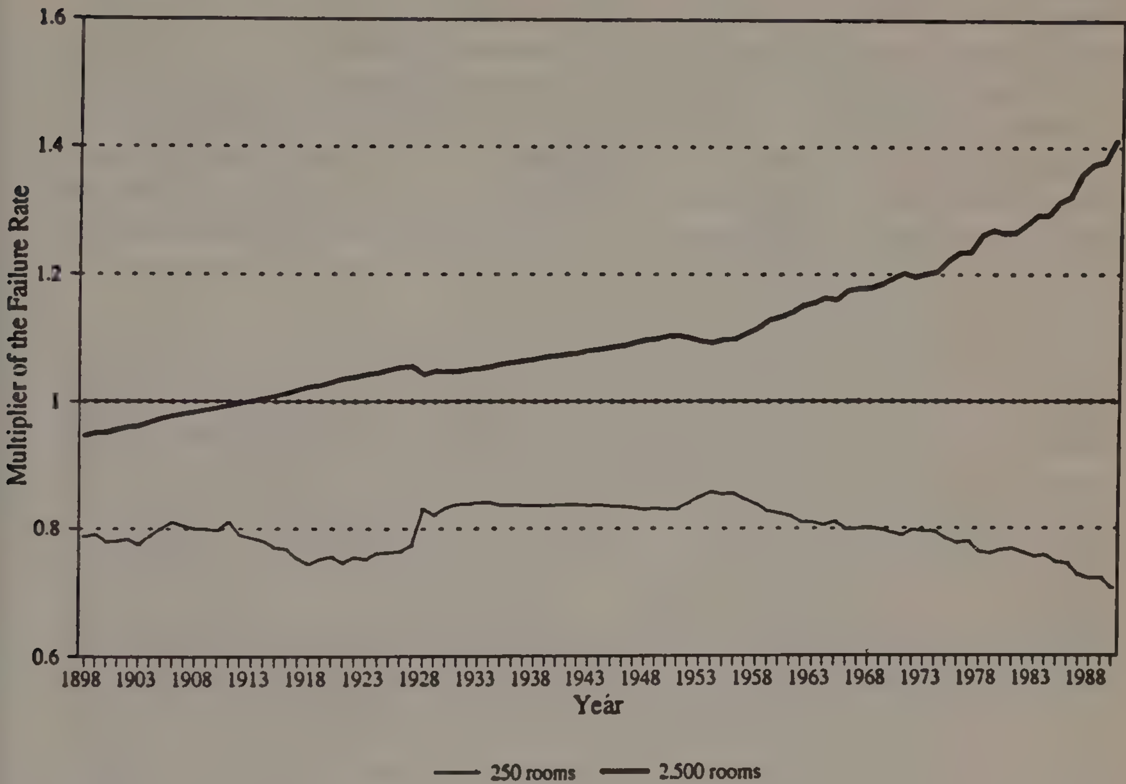
FIGURE 5: Size-Localized Effects (Model 4)



the industry age × $D_{it}^{size/2}$ interaction is significant and negative. Given that the main effect of $D_{it}^{size/2}$ on the failure rate is not significant, the interaction effect suggests that the intensity of size-localized competition was initially low but increased over the course of the Manhattan hotel industry's history, supporting hypothesis 3b. Figure 5 illustrates this temporal variation in the effect of size-localized competition graphically. This figure gives a multiplier that shows the effect of reducing $D_{it}^{size/2}$ by the average Euclidean distance to another hotel in each year. As Figure 5 shows, the multiplier for increasing size-localized competition by one hotel of average size similarity varies between 1.02 and 1.03 from 1898 to 1928, drops with the population's rapid growth, and then increases continuously, reaching nearly 1.05 by 1990. Thus, like mass-dependent competition, the importance of size-localized competition increased over the course of the industry's history.

Finally, model 5 introduces the industry age × size × market concentration interaction.[11] The two additional two-way interactions included in the model are required to obtain an unbiased estimate for the three-way interaction of interest (Cohen & Cohen 1983). Again, adding these interaction terms improves the fit of the model significantly ($\Delta\chi^2 = 20.02$, df = 3, $p < .05$). Supporting hypothesis 4b, the coefficient for the industry age × size × market concentration interaction is significant and positive. However, contrary to resource partitioning theory, the main effect of market concentration remains significant and positive and the size

FIGURE 6: Market Concentration Effects (Model 5)



× market concentration interaction remains significant and negative. The industry age × market concentration interaction is also significant and negative. The complex implications of these interactions are depicted graphically in Figure 6. This figure gives multipliers that show how increasing market concentration in each year by 10% affects the failure rate for hotels with 250 and 2,500 rooms. Three features of Figure 6 are noteworthy. First, the multiplier is always higher for the 2,500-room hotel than for the 250-room hotel. Second, the multiplier for the 250-room hotel is always less than 1, indicating that increases in concentration always lower its failure rate. Third, the multiplier for the 2,500-room hotel increases over the observation period from 0.95 to 1.4, while the multiplier for the 250-room hotel falls from 0.8 to 0.7. Thus, the positive industry age × size × market concentration interaction dominates the negative size × market concentration interaction, resulting in an overall effect that is consistent with the predictions of resource partitioning theory (hypothesis 4a) and hypothesis 4b.

The organizational and environmental control variables also had several significant influences on hotel failure rates. In contrast to the commonly observed age-dependent decline in organizational failure rates, the significant positive estimate for age indicates that hotel failure rates increased with age. Thus, a hotel is an expensive physical asset that, if not properly maintained, may be subject to obsolescence (Baum & Mezias 1992). However, consistent with

prior work, hotel failure rates did exhibit negative size dependence. Failure rates were also lower among luxury hotels than among economy hotels. The estimate for the left-censored dummy variable is insignificant in all models. Thus, although hotels founded before the observation period were already in a sense "survivors," they did not exhibit a survival advantage.

Among the environmental factors, increases in potential demand (i.e., the number of visitors to New York City) and the GNP growth rate both lowered the failure rate. The failure rate also declined as resources released by hotel failures in the prior year enhanced the viability of surviving hotels (Delacroix, Swaminathan & Solt 1989) but was unrelated to the number of prior foundings.

Discussion and Conclusion

This article began by noting that density-dependence theory alone cannot account for the widely observed pattern of proliferation and concentration in organizational populations. Although density dependence provides an explanation for the initial proliferation to an equilibrium level, it does not explain the subsequent concentration because it does not permit asymmetric competition that allows some population members to become dominant. Prompted by this basic limitation of density-dependence theory and research, I attempted to contribute to an explanation for this evolutionary tendency of organizational populations in two ways. First, I combined density dependence with complementary models of ecological competition that incorporate size-based competitive asymmetries among the members of a population. Second, based on the original statement of density-dependence theory in organizational ecology (Brittain & Freeman 1980), I advanced hypotheses that predict changes in the strengths of competitive processes over the course of a population's history. The findings show how incorporating time variation in multiple competitive processes can provide a more detailed understanding of the competitive dynamics and evolution of an organizational population.

My analysis of the Manhattan hotel industry demonstrates that hotels of different sizes generate and experience different strengths of competition and that the strengths of density- and size-based competitive processes change over time. Although there is evidence that four distinct competitive processes influenced the Manhattan hotel industry simultaneously, I draw attention to the results establishing the changing basis of competition, which provide new insights into the propensity of older organizational populations to become concentrated. Supporting the idea that, over time, resource partitioning and organizational differentiation reduce the number of direct competitors each organization faces within a population, the competitive effect of population density weakened over the observation period (hypothesis 1b). However, consistent with the idea that as competition on the basis of efficiency intensifies over time, larger organizations become stronger competitors than their smaller rivals (Brittain & Freeman 1980) and mass-dependent competition intensifies (hypothesis 2b). The results also support the prediction that similarly sized organizations become more intense competitors over time (hypothesis 3b). This is compatible with the idea that as resource partitioning proceeds, competition

among large organizations to occupy central markets intensifies, leaving smaller organizations to compete with each other for more specialized resource opportunities. Lastly, and also consistent with the prediction that large organizations focus increasingly on competing with each other as resource partitioning proceeds, the harmful effect of market concentration on the survival of large organizations increased over time (hypothesis 4b).

Overall, the findings extend empirical evidence linking competitive processes to population dynamics by demonstrating that, while density dependence was vital to the initial proliferation of the Manhattan hotel industry, size-based competitive asymmetries became increasingly important over time, helping to account for the industry's later decline in numbers and concentration. Thus, our understanding of the evolution of the industry is enhanced by combining ideas from four models of ecological competition and by permitting the effects of competitive processes specified by each model to vary over the industry's history instead of restricting the effects of these variables to be time-invariant (Isaac & Griffin 1989).

Although the basic findings of this study the were broadly consistent with theoretical expectations, the mutualistic effect of the presence of large hotels early in the observation period provides an interesting exception. Several factors may account for this early mutualistic effect of large hotels. While large hotels may generally have been stronger competitors, their political and economic strength may also have facilitated the early growth of the industry. Manhattan's early "grand hotels" were a new phenomenon, unlike the inns or taverns typical of the time. Visitors to the city told wondrous stories of Manhattan's grand hotels; they were cities within a city, completely self-contained "Palaces of the People, where any man, with a few extra dollars to spend, could live like a king, at least for a few days" (Batterberry & Batterberry 1973:80). Thus, the initial emergence of large hotels may have contributed to the institutional legitimacy of the hotel industry in Manhattan and may also have helped stabilize the resource environment by helping to attract a regular flow of business and tourist travelers to the New York City area. Small hotels may have benefited from the presence of large hotels in a more direct way if proximity to large hotels was a indication to travelers of the desirability of their location. Moreover, smaller and newly established hotels may have been able to improve their survival chances by drawing on the strengths of the industry's early large hotels, for example, by servicing their overflow clients.

As the industry grew, and the basis of competition shifted toward an emphasis on cost and efficiency, the increasing competitive strength of the industry's large members made it increasingly difficult for smaller participants to compete effectively, overwhelming the mutualistic effects. However, smaller hotels were somewhat sheltered from their larger rivals as competition among them to occupy central markets intensified, leaving smaller hotels to compete among themselves for peripheral resources. Thus, large hotels may have played an important role both in the initial expansion of the industry and also in its later decline in size as a smaller number of larger hotels displaced the industry's smaller participants.

Several implications and directions for future research follow from the results of this study. First, as with any other single-industry study, the results

may reflect industry-specific factors. Future replications of this study in other populations are needed to help establish the generalizability and broader significance of the findings. For example, although this study provides evidence that the basis of competition changes in the Manhattan hotel industry, future research is needed to examine predictions regarding the conditions that give rise to these changes. Are there industries in which the basis of competition changes in different ways or not at all? One interesting possibility is Barnett and Amburgey's (1990) suggestion that the competitive reach of large organizations is affected by the heterogeneity of the population's niche. In heterogeneous niches the competitive reach of a large organization is predicted to be localized to its segment(s) of the niche, increasing the potential for size-localized competition. By comparison, in homogeneous niches the competitive reach of large organizations is extensive, increasing the potential for mass-dependent competition. In moderately heterogeneous niches, such as the Manhattan hotel industry, the potential for both mass-dependent and size-localized competition exists. Research examining the effects of niche heterogeneity may therefore improve our understanding of factors underlying changes in the basis of competition.

Second, while this research concentrates on variations in the intensity of competitive relationships that depend on differences in organizational size, theory in organizational ecology suggests that similarity more generally (e.g., similarity in targeted clients, goods and services produced, and technologies employed) is a primary determinant of the intensity of competition between organizations (e.g., Aldrich 1979; Baum & Singh 1994a, 1994b; Hannan & Freeman 1977, 1989; McPherson 1983).[12] Among hotels, for example, resource requirements, and thus the intensity of competition, vary not only with similarity in size, but also with similarity in price and location (Baum & Mezias 1992; Lant & Baum 1995). Traditionally, hotels have positioned themselves in one of three service-level categories oriented toward well-defined market segments: luxury, midprice, or economy. This results in patterns of resource use that are specialized to regions of the price distribution, intensifying competition between similarly priced hotels. Hotels also compete more intensely within relatively local market areas defined by their proximity to points of tourist interest or business activity from which the demand for hotel services is derived. Thus, competition is most intense between establishments of similar class or market position within relatively compact and well-defined geographic areas or hotel districts (Baum & Mezias 1992; Lant & Baum 1995). Research also shows that competitive intensity varies with similarity in the characteristics of affiliated members of voluntary associations (McPherson 1983, 1990) and similarity in the ages of children enrolled in day-care centers (Baum & Singh 1994a, 1994b). Future research that considers organizational differentiation more broadly may therefore improve our understanding of factors influencing competitive intensity and evolution of organizational populations.

Third, although this study has drawn some initial links among ecological models of competition by examining how their strengths vary over an organizational population's history, clearly the relationships among of these models need to be explored more fully. Ecological theory can be enriched by integrating further ideas from multiple models of population dynamics (Singh 1993). For

example, this study focused on how changes in the basis of competition over a population's history influence patterns of competition among organizations of different sizes. However, Brittain and Freeman's (1980) density-dependence model and Carroll's (1985) resource partitioning model both have broader implications for the changing mix and relative success of strategies within a population of organizations over time. As described earlier, Carroll focuses on specialist and generalist strategies and Brittain and Freeman focus on *r*- and *K*-strategists. While specialists and *r*-strategists are predicted to thrive in emergent population, generalists and *K*-strategists are predicted to dominate older populations. However, size is only one organizational feature related to the differences between specialists and generalists and *r*- and *K*-strategists. Notably, these ecological categories encompass concepts of market breadth and market aggressiveness, which are central to classifications of competitive strategy in the strategic management literature (Romanelli 1989). Consequently, future studies that consider organizational differences in strategic orientation may provide a bridge between research in organizational ecology and the field of strategic management, which is also interested in explaining why some strategies lead to competitive superiority while others do not.

Finally, this study brings into question past research that tests for density dependence in populations that have evolved beyond their peak density. Systematic changes in the sizes and strategies of organizations over a population's history suggest that the adequacy of population density as a proxy for competition is diminished over a population's history. Therefore, I modeled density dependence together with three size-based models of competition and permitted time variation in the density-squared effect to account for these changes and reduce specification bias due to unobserved heterogeneity in the density estimates (Tuma & Hannan 1984). Furthermore, when density is equally low early and late in a population's history, the legitimacy account of the coefficient for the linear density term is unclear because late low density has no meaning in density-dependence theory. Therefore, since population density is equally low early and late in the Manhattan hotel industry's history (see Figure 1), I separated the effects of early and late low-density conditions empirically. Notably, after accounting for the effects of size-based competition and permitting time variation in density-dependence effects, density-dependent predictions of legitimation and competition were supported. This increases my confidence in the density-dependence model, but it also raises questions about the veracity of past research findings — both supportive and unsupportive. However, it is not the purpose of this article to either criticize or defend density-dependence theory. It is to challenge the inertia in organizational ecology research on competition and industry evolution by pointing out some key limitations of density dependence, while, at the same time, showing that these limitations can be overcome and new insights gained by adopting a broader, multitheoretical approach that combines density dependence with other well-established ecological models of competition.

This research is the first to examine the changing basis of competition in organizational populations. The findings of this study demonstrate the potential significance of such changes for a basic evolutionary property of organizational populations: the tendency of older populations to become concentrated. Given

the evidence linking temporal changes in the basis of competition to population evolution, I believe future research that attends to the historical as well as diverse ecological processes of organizational populations will provide new insights into organizational evolution.

Notes

1. In more technical terms, when tested in populations that have evolved beyond their peak density, population density estimates suffer from specification bias due to unobserved heterogeneity (Tuma & Hannan 1984). This occurs when one or more independent variables that influence the dependent variable are omitted from the analysis.

2. Similar arguments are made in models of spatial competition in economics. These models predict that competition among products in a bounded characteristics space leads to product differentiation, which leads, in turn, to a decrease in the intensity of product competition (e.g., Swann 1985, 1990).

3. Although the period of rapid growth in the late 1920s coincides roughly with the addition of the *Yellow Pages* data source, it is not attributable to the addition of this source. This growth, which accompanied the rapid expansion of Manhattan during this time, is reflected in the *Red Book* as well.

4. This is a departure from Barnett and Amburgey's (1990) measure of population mass, the total number of telephones. However, as Winter (1990) points out, Barnett and Amburgey's variable is arguably a measure of the demand served by other organizations rather than their capacity.

5. The assumption that a focal hotel competes only with other hotels within a certain range of its own size is consistent with the earlier analysis of these data by Baum and Mezias (1992).

6. Baum and Mezias (1992:597-98) discuss the robustness of $D_{it}^{size \pm \mu}$ to changes in the value of μ. They also discuss and model the effects of localized competition in terms of hotels' geographic location and price (pp. 589-91). Adding these forms of localized competition to the models reported below does not change the substantive findings.

7. This starting date was used for industry age for comparability. Results reported below do not differ substantively if the industry age clock is started in the first observation year.

8. Hannan and Carroll (1995) use a similar approach. However, they do not include main effects for industry age or industry age squared in their interaction models. This biases their estimates, making interpretation of their findings problematic (Cohen & Cohen 1983).

9. Although membership in a hotel chain may influence the survival of individual hotels, chain membership data are not available in the *Red Book* and so could not be controlled in this study.

10. These two variables were constructed using information contained in the *Historical Statistics of the United States, Colonial Times to 1970* (U.S. Department of Commerce, Bureau of the Census, 1975), *Historical Abstracts of the United States* (U.S. Department of Commerce, Bureau of the Census, 1970-1990), and *Port of New York and New Jersey Authority* annual reports (1930-1990).

11. The implications of models 2-4 in Figures 3-6 are not altered substantively using estimates from model 5.

12. Similar views are taken in strategic (e.g., Porter 1980) and economic (e.g., Scherer & Ross 1990; Tirole 1988) theories of competition.

References

Aldrich, Howard E. 1979. *Organizations and Environments*. Prentice-Hall.

Aldrich, Howard E., and Ellen R. Auster. 1986. "Even Dwarfs Started Small: Liabilities of Age and Size and Their Strategic Implications." *Research in Organizational Behavior* 8:165-98.

Bain, Joe S. 1956. *Barriers to New Competition.* Harvard University Press.

Banaszak-Holl, Jane. 1992. "Historical Trends in Rates of Manhattan Bank Mergers, Acquisitions and Failures." Unpublished manuscript, Center for Health Care Research, Brown University.

Barnett, William P., and Terry L. Amburgey. 1990. "Do Larger Organizations Generate Stronger Competition?" Pp. 78-102 in *Organizational Evolution: New Directions,* edited by Jitendra V. Singh. Sage.

Batterberry, Michael, and Ariane Batterberry. 1973. *On the Town in New York from 1776 to the Present.* Scribner's.

Baum, Joel A.C., and Stephen J. Mezias. 1992. "Localized Competition and the Dynamics of Organizational Failure in the Manhattan Hotel Industry." *Administrative Science Quarterly* 37:483-501.

Baum, Joel A.C., and Jitendra V. Singh. 1994a. "Organizational Niche Overlap and the Dynamics of Organizational Mortality." *American Journal of Sociology* 100:346-80.

____. 1994b. "Organizational Niche Overlap and the Dynamics of Organizational Founding." *Organization Science* 5:00-00.

Brittain, Jack W., and John Freeman. 1980. "Organizational Proliferation and Density-Dependent Selection." Pp. 291-338 in *Organizational Life Cycles,* edited by John Kimberly and Raymond Miles. Jossey-Bass.

Carpenter, G.S., L.G. Cooper, D.M. Hanssens, and D.F. Midgley. 1988. "Modeling Asymmetric Competition." *Marketing Science* 7:393-411.

Carroll, Glenn R. 1985. "Concentration and Specialization: Dynamics of Niche Width in Populations of Organizations." *American Journal of Sociology* 90:1262-83.

Carroll, Glenn R., and Michael T. Hannan. 1989. "Density Dependence in the Evolution of Populations of Newspaper Organizations." *American Sociological Review* 54:524-41.

Carroll, Glenn R., and Anand Swaminathan. 1992. "The Organizational Ecology of Strategic Groups in the American Brewing Industry from 1975 to 1990." *Industrial and Corporate Change* 1:65-97.

Caves, Richard E., and Michael E. Porter. 1977. "From Entry Barriers to Mobility Barriers." *Quarterly Journal of Economics* 90:241-61.

Cohen, Jacob, and Patricia Cohen. 1983. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* 2d ed. Erlbaum.

Delacroix, Jacques, Anand Swaminathan, and Michael E. Solt. 1989. "Density Dependence versus Population Dynamics: An Ecological Study of Failings in the California Wine Industry." *American Sociological Review* 54:245-62.

Edwards, Corwin D. 1955. "Conglomerate Bigness as a Source of Power." Pp. 331-59 in *Business Concentration and Public Power,* edited by the National Bureau of Economic Research. Princeton University Press.

Hannan, Michael T. 1986. "Competitive and Institutional Processes in Organizational Ecology." Technical Report 86-13, Dept. of Sociology, Cornell University.

Hannan, Michael T., and Glenn R. Carroll. 1992. *Dynamics of Organizational Populations: Density, Competition, and Legitimation.* Oxford University Press.

____. 1995. "Theory Building and Cheap Talk about Legitimation: Reply to Baum and Powell." *American Sociological Review* 60:539-44.

Hannan, Michael T., and John Freeman. 1977. "The Population Ecology of Organizations." *American Journal of Sociology* 83:929-84.

____. 1989. *Organizational Ecology.* Harvard University Press.

Hannan, Michael T., James Ranger-Moore, and Jane Banaszak-Holl. 1990. "Competition and the Evolution of Organizational Size Distributions." Pp. 246-68 in *Organizational Evolution: New Directions,* edited by Jitendra V. Singh. Sage.

Hawley, Amos H. 1950. *Human Ecology: A Theory of Community Structure*. Ronald Press.

Ijiri, Yuji, and Herbert A. Simon. 1977. *Skew Distributions and the Sizes of Business Firms*. North Holland.

Isaac, Larry W., and Larry J. Griffin. 1989. "Ahistoricism in Time-Series Analyses Of Historical Process: Critique, Redirection, and Illustrations from U.S. Labor History." *American Sociological Review* 54:873-90.

Lant, Theresa K., and Joel A.C. Baum. 1995. "Cognitive Sources of Socially Constructed Competitive Groups: Examples from the Manhattan Hotel Industry." Pp. 31-47 in *The Institutional Construction of Organizations*, edited by W. Richard Scott and Soren Christensen. Sage.

McPherson, J. Miller. 1983. "An Ecology of Affiliation." *American Sociological Review* 48:519-32.

____. 1990. "The Evolution of Communities of Voluntary Organizations." Pp. 224-45 in *Organizational Evolution: New Directions*, edited by Jitendra V. Singh. Sage.

Meyer, Marshall W. 1990. "Notes of a Skeptic: From Organizational Ecology to Organizational Evolution." Pp. 298-314 in *Organizational Evolution: New Directions*, edited by Jitendra V. Singh. Sage.

Pfeffer, Jeffrey, and Gerald R. Salancik. 1978. *External Control of Organizations*. Harper & Row.

Porter, Michael E. 1980. *Competitive Strategy*. Free Press.

Romanelli, Elaine. 1989. "Environments and Strategies of Organizational Start-up: Effects on Early Survival." *Administrative Science Quarterly* 34:369-87.

Scott, W. Richard. 1987. *Organizations: Rational, Natural and Open Systems*. 2d ed. Prentice-Hall.

Scherer, F.M., and David Ross. 1990. *Industrial Market Structure and Economic Performance*. 3d ed. Houghton Mifflin.

Singh, Jitendra V. 1993. "Density Dependence Theory — Current Issues, Future Promise." Review Essay on *Dynamics of Organizational Populations* by Michael T. Hannan and Glenn R. Carroll." *American Journal of Sociology* 99:464-73.

Singh, Jitendra V., and Charles J. Lumsden. 1990. "Theory and Research in Organizational Ecology." *Annual Review of Sociology* 16:161-95.

Starbuck, William H. 1965. "Organizational Growth and Development." Pp. 451-533 in *Handbook of Organizations*, edited by James G. March. Rand McNally.

Swann, G.M.P. 1985. "Product Competition in Microprocessors." *Journal of Industrial Economics* 34:33-53.

____. 1990. "Product Competition and the Dimensions of Product Space." *International Journal of Industrial Organization* 8:281-95.

Thompson, James D. 1967. *Organizations in Action*. Academic Press.

Tirole, Jean. 1988. *The Theory of Industrial Organization*. MIT Press.

Tuma, Nancy B. 1980. *Invoking Rate*. 2d ed. SRI.

Tuma, Nancy B., and Michael T. Hannan. 1984. *Social Dynamics: Models and Methods*. Academic Press.

Utterback James M., and Fernando F. Suárez. 1993. "Innovation, Competition, and Industry Structure." *Policy Studies* 22:1-21.

Wholey, Douglas R., Jon B. Christianson, and Susan M. Sanchez. 1992. "Organization Size and Failure among Health Maintenance Organizations." *American Sociological Review* 57:829-42.

Winter, Sidney G. 1990. "Survival, Selection, and Inheritance in Evolutionary Theories of Organization." Pp. 269-97 in *Organizational Evolution: New Directions*, edited by Jitendra V. Singh. Sage.

Wyckoff, D. Daryl, and W. Earl Sasser. 1981. *The U.S. Lodging Industry*. D.C. Heath.

Interracial Contact and Black Racial Attitudes: The Contact Hypothesis and Selectivity Bias*

DANIEL A. POWERS, *University of Texas at Austin*
CHRISTOPHER G. ELLISON, *University of Texas at Austin*

Abstract

One of the most serious criticisms of research on the contact hypothesis is the contention that selection bias operates to promote interaction between whites and blacks who are already relatively unprejudiced toward one another. Accordingly, attempts to infer the effect of interracial contact on racial attitudes must recognize and correct for this potential source of bias. Endogenous switching regression models are used to estimate the effect of close interracial friendship on selected racial attitude variables while accounting for possible selectivity bias. Each model implies a distinct assumption about the process generating interracial contact and racial attitudes. Using data from the National Survey of Black Americans, we find no evidence of sample selection bias in estimating the effect of close interracial contact on black racial attitudes. Therefore, we cannot reject the conventional models used in research on the contact hypothesis that treat interracial contact as an exogenous variable affecting racial attitudes. However, our results show the usefulness of a switching regressions approach to reveal the contingent nature of the contact hypothesis.

For at least four decades, social scientists have debated the value of the contact hypothesis for the study of racial and ethnic relations (Allport 1954; Jackman & Crane 1986; Sigelman & Welch 1993; Stephan 1987). This hypothesis holds that contact, particularly close and sustained contact, with members of different racial and ethnic groups promotes positive, tolerant attitudes toward those groups. Proponents of the contact hypothesis maintain that interracial contact provides direct information — frequently accurate and favorable information — on the values, lifestyles, and experiences of other groups. This positive firsthand information may be generalized into a positive perception of the group(s) as a whole, thus permitting individuals to counter unfavorable racial and ethnic stereotypes. Commentators have also presented more conditional versions of the

* *Data were made available by the Interuniversity Consortium for Political and Social Research, Ann Arbor. We are grateful to Arthur Sakamoto and the anonymous reviewers for helpful comments on an earlier version. However, the authors are responsible for the analyses and interpretations presented here. Direct correspondence to Daniel A. Powers, Department of Sociology, 336 Burdine, University of Texas at Austin, Austin, TX 78712-1088.*

contact hypothesis, suggesting that contact may promote positive racial and ethnic attitudes only under ideal conditions — for instance, when the contact involves persons of equal status, when it takes place under cooperative conditions, and when it enjoys the active support of powerful authorities (Allport 1954; Desforges et al. 1991).

To date, research on the contact hypothesis has been plagued by three critical limitations. First, most empirical investigations over the years have been conducted in specific institutional contexts (e.g., interracial housing projects, desegregated schools, the military) or in laboratories (Brown & Albee 1966; Butler & Wilson 1978; Desforges et al. 1991; Wilner, Walkley & Cook 1955). However, most interracial contact does not occur in these specialized, highly monitored settings. Thus, while these studies have contributed to our understanding of intergroup relations, until recently they have left open the question of whether interracial contact results in positive racial attitudes in the general population. Three recent studies using general population samples have reached divergent conclusions about the usefulness of the contact hypothesis for the study of race relations (Ellison & Powers 1994; Jackman & Crane 1986; Sigelman & Welch 1993).

Second, research on the contact hypothesis has focused primarily on white racial attitudes, giving short shrift to the views of African Americans. The handful of exceptions to this general pattern over the years have yielded discrepant findings (Ford 1973; McDowell 1971; Robinson & Preston 1976; Works 1961). Because many African Americans are forced to encounter whites frequently in public settings, due to the demographic imbalance between blacks and whites, and because such contact often takes place under competitive and even hostile circumstances, it appears that casual interracial contact is not necessarily associated with positive racial attitudes among African Americans. However, more intimate interracial contacts, such as close white friendships, seem to promote positive racial attitudes among African Americans (Ellison & Powers 1994; Sigelman & Welch 1993). Given mounting evidence that the contact hypothesis works differently for blacks than for whites, additional research on the implications of interracial contact for the racial attitudes of African Americans seems warranted.

Third, several recent studies have raised important questions on the causal assumptions of the contact hypothesis (Jackman & Crane 1986; Sigelman & Welch 1993). For instance, some have suggested that the positive association between contact and positive attitudes reported in cross-sectional studies may reflect primarily a selection effect: Initially tolerant attitudes may lead individuals to engage in, or even to seek out, interracial contacts, while less tolerant persons eschew such contacts. If this interpretation is accurate, then research on the contact hypothesis for the past four decades has been fundamentally misguided, and further inquiry on the consequences of interracial contact holds little promise for social scientists or policymakers. The urgency of this question has been underscored by Sigelman and Welch (1993), who concluded their recent analysis by noting that "our findings pinpoint the issue [of causal order] as a live one, the resolution of which should be a priority focus in future research on blacks' and whites' racial attitudes" (793).

This study contributes to the literature on the contact hypothesis in several ways. In the absence of large-scale panel data on interracial contacts and racial attitudes, we propose endogenous switching regression models to examine issues of causality and selectivity that have dogged recent research on the contact hypothesis. After elaborating the rationale for this approach, we then test several sets of hypotheses concerning black racial attitudes, using data from a large ($N = 2{,}107$), nationally representative cross section of African Americans.

The Selection Problem in Research on the Contact Hypothesis

One of the most serious criticisms of research on the contact hypothesis is the contention that selection bias operates to promote interaction between whites and blacks who are already relatively unprejudiced toward one another. Most statistical models of the relationship between contact and attitudes assume that contact is an independent (exogenous) variable influencing racial attitudes. This implies that, net of observed covariates, contact and attitudes are statistically independent. However, it may be that initially tolerant attitudes lead persons to engage in or even seek out interracial contact.

Models based on other assumptions about the process generating interracial contact and racial attitudes may yield different estimates of the effects of key variables. If contact is viewed as an exogenous variable in a model predicting racial attitudes, the estimated effect of contact is simply the difference between the average attitude measure for those who have experienced contact and those who have not. However, if observed attitude differences between those who have experienced contact and those who have not are due to predispositions toward racial tolerance rather than interracial contact per se, then the effect of interracial contact is no longer simply a difference in average attitude measures. The estimated effect of contact and other key variables in regression models of racial attitudes will be inconsistent as a result of a selection bias that sorts people nonrandomly into situations of contact with members of other racial and ethnic groups.

A selection model interpretation offers a more fruitful approach for the analysis of interracial contact and racial attitudes by altering the causal sequence surrounding attitude change implicit in most contact research. In such models, common unmeasured factors would affect both the probability of engaging in interracial contact and the strength of particular racial attitudes.

Data and Measures

Data from the National Survey of Black Americans (NSBA) (Jackson, Tucker & Gurin 1987) are used to test models alternative to the contact hypothesis. The NSBA is a survey conducted by the Survey Research Center (SRC) at the University of Michigan during 1979-80. The sampling and interview procedures, which are described elsewhere in more detail (see Jackson 1991), yielded a sample of 2,107 adults.

DEPENDENT VARIABLES

In this article we seek to explain the observed association between interracial contact and racial attitudes while controlling for observed covariates and taking account of possible selection bias. We take a closer look at racial attitude variables measuring (1) the amount of skepticism that blacks feel toward whites, and (2) the degree of blacks' opposition toward interracial dating. These variables are standard indicators that have appeared in the literature on racial attitudes (see Schuman & Hatchett 1974; Schuman, Steeh & Bobo 1988).

First, we consider blacks' *skepticism about the motivations of whites.* This information is obtained from responses to the question; "On the whole, do you think most White people want to see blacks get a better break, or do they want to keep blacks down, or don't they care one way or another?" Approximately 22% of NSBA respondents supplied the most favorable response, while an additional 35.9% answered that whites are indifferent to the situation of blacks. The remaining 42.1% felt that whites seek to oppress blacks.

Second, we consider blacks' *disapproval of interracial dating,* measured in terms of the strength of the respondent's agreement with the following two statements: (1) Black men should not date white women; and (2) Black women should not date white men ($r = .80$). These items range from "strongly disagree" (coded 1) to "strongly agree" (coded 4). Our indicator is the mean response to these two items, which range from 1 to 4 with a mean of 2.145 and a standard deviation of 1.41.

To assess the effect of interracial contact on racial attitudes, we assume that blacks who have close white friends occupy a distinct position or status from blacks who do not. The availability of *close white friends* is measured by the following dichotomous item: "Do you know any White person who you think of as a good friend — that is, someone to whom you can say what you really think?" Approximately 57% of the NSBA respondents reported having at least one such white friend.

INDEPENDENT VARIABLES

We control for the influence of background factors that are established predictors of racial attitudes (see Demo & Hughes 1990; Ellison 1991): *age* (in years); *education* (8-point summary scale); gender (1 = *female*); *family income* (17-point summary scale); *urban* vs. rural residence (1 = resident of self-representing urban area); *South* (1 = resident of southern state, according to the U.S. Bureau of the Census classification).

In addition to these control measures, we include two additional racial contact variables that influence attitudes and interracial friendship formation. *Childhood contact* refers to casual contact with whites that occurred earlier in life. This is measured in terms of the racial composition of the respondent's elementary school, junior high school, high school, college, and childhood neighborhood. *Adult contact* refers to the level of casual interracial contact experienced by the respondent at the time of the NSBA interview. This is measured in terms of the interracial composition of the respondent's current neighborhood, church or place of worship, and workplace. The responses on

these items range from "all blacks" (coded 1) to "almost all whites" (coded 5). Our measure of each contact variable is the mean score on the corresponding set of items.

Although it is reasonable to expect that, other things being equal, persons who are immersed in interracial settings regularly will actually experience more interpersonal contact with individuals of other racial and ethnic backgrounds, we acknowledge that these indicators cannot measure actual contact precisely. Nevertheless, these indicators are identical to those used in several recent studies of racial attitudes among whites and blacks (e.g., Ellison & Powers 1994; Sigelman & Welch 1993). Thus, our use of these measures maximizes continuity with previous research in this area.

ENDOGENOUS SWITCHING REGRESSION MODELS FOR CAUSAL EFFECTS

The selection problem of contact research is closely related to the problem of estimating the effects of statuses, social positions, and treatment effects from sample surveys (Blalock 1964; Winship & Mare 1992). White friendship may be regarded as a treatment variable representing two groups, "no white friends" (coded 0) and "white friends" (coded 1), into which blacks are assigned, possibly nonrandomly. The extent to which assignment to each of these groups is nonrandom can be investigated with endogenous switching models (Mare & Winship 1988; Winship & Mare 1992).[1] As Winship and Mare (1992) note, these models are "suitable for assessing the effects of a social classification . . . on a consequence of membership in this classification" (338).

Following the conventional notation for switching regression models (see Mare & Winship 1988; Winship & Mare 1992), we assume that NSBA respondents are drawn from a population of blacks, each of whom is characterized by values for the variables y_1, y_0, z, x. Each person is characterized by two hypothetical attitudinal outcomes: y_1 denotes the attitudinal outcome that would have been observed in the black population among those having close white friends; and y_0 denotes the outcome that would have resulted among those *not* having close white friends. Let $z = 1$ if a respondent reported having white friends, and 0 otherwise. Because either attitudinal outcome y_1 or attitudinal outcome y_0 is realized, it follows that y_1 is latent (or missing) when y_0 is observed ($z = 0$), and y_0 is latent (or missing) when y_1 is observed ($z = 1$). The *realized* outcome $y = zy_1 + (1 - z)y_0$.[2]

Attitudinal outcomes (y_1 and y_0) and white friendship status (z) depend on a collection of predetermined exogenous sociodemographic and control covariates x_j ($j = 1, \dots J$), where $x_1 = 1$ for all individuals to allow for constant terms. With no loss of generality, y_1, y_0, and z may depend on common or unique subsets of x_j; some covariates may affect racial attitudes, others may influence the formation of close friendships with whites, and some may affect both attitudes and friendship formation. Following Mare and Winship (1988), the system of equations for the ith individual can be written as

$$y_{1i} = \sum_j \beta_{1j} x_{ji} + \epsilon_{1i} \qquad \text{if } z_i = 1 \tag{1a}$$

$$y_{0i} = \sum_j \beta_{0j} x_{ji} + \epsilon_{0i} \quad \text{if } z_i = 0 \tag{1b}$$

$$z_i = 1 \text{ if } \sum_j \gamma_j x_{ji} + \eta_1 y_{1i} + \eta_0 y_{0i} + \nu_i > 0 \tag{1c}$$
$$= 0 \text{ otherwise}$$

where β_{kj} ($k = 0,1$) denotes the regression effect of x_j on the attitude measures, γ_j denotes the corresponding effect of x_j on white friendship, η_1 and η_0 are the effects of racial attitudes on white friendships, and $x_{1i} = 1$ for all observations to allow for equation intercepts. It is worth noting that in restricted versions of this model some x_j may be excluded from certain equations above and some effects on attitudinal outcomes may be the same for persons in either white friendship position.

We assume that errors are independently distributed, trivariate normal with a mean of zero and variances of σ_0, σ_1, and 1, υ is uncorrelated (conditional on x) with the remaining equation errors (ε_1 and ε_0), and all x_j are uncorrelated with the equation disturbances. The covariance σ_{01} is fixed at zero. This term is not identified, as each person realizes either y_1 or y_0.

A reduced form of the z equation is useful in the subsequent applications.

$$z_i = \sum_j \pi_j x_{ji} + \epsilon_{zi} \tag{2}$$

where

$$\pi_j = \gamma_j + \eta_1 \beta_{1j} + \eta_0 \beta_{0j}, \text{ and } \epsilon_{zi} = \eta_1 \epsilon_{1i} + \eta_0 \epsilon_{0i} + \nu_i.$$

In general, the error term of the reduced form equation, ε_z, will be correlated with ε_1 and ε_0.[3]

We make use of continuous and dichotomous measures of racial attitudes. For the skepticism measures, we focus separately on the most positive and most negative racial attitudes. We noted earlier that approximately 22% of the NSBA sample expressed the most positive attitude that "Whites wanted to see Blacks get a better break," whereas approximately 35.9% of the NSBA sample expressed the negative racial attitude that "Whites want to keep Blacks down."

Each response from the skepticism measure can be treated as a dichotomous dependent variable. Those respondents who express another attitude than the positive are neutral or have a negative attitude. Similarly, those respondents expressing an attitude differing from the most negative attitude are either neutral or have positive attitudes. In the case of the positive racial attitude, $y_1 = 1$ if a black person with close white friends expressed the attitude that whites seek to give blacks a break, and zero otherwise. Likewise, $y_0 = 1$ if a black person without close white friends believed that whites seek to give blacks a break, and zero otherwise. Similar expressions obtain for the negative racial attitude. Each equation in the system above would be estimated as a probit regression subject to the standard normalization, $\sigma_1 = \sigma_0 = 1$ with error correlations ρ_{1z} and ρ_{0z}.

The interracial dating measure is treated as continuous. Here, y_1 represents the degree of disapproval of interracial dating expressed by a black American with close white friends, and y_0 is the same attitude of a black American without close white friends. The two attitude equations would be estimated as OLS regressions with error covariances σ_{1z} and σ_{0z}.

Models and Their Implications

Several competing models can be estimated, each of which tests assumptions implicit in contact research by placing restrictions on the parameters in equations 1 and 2. These restrictions are outlined in more detail in the Appendix.

MODEL 1: THE GENERAL MODEL

The general model imposes the fewest restrictions. The magnitudes and signs of the estimated error covariances (correlations) provide information about the degree and direction of nonrandom selection into one or another of the white friendship statuses. For example, $\sigma_{1z}>0$ (or $\rho_{1z}>0$, if equations 1a and 1b are probit equations) provides evidence of positive selection of persons into the "white friends" position, whereas σ_{1z} (ρ_{1z}) <0 implies negative selection into the "white friends" status. Similarly, σ_{0z} (ρ_{0z}) <0 provides evidence of positive selection of persons into the "no close white friends" position. Thus, net of observed covariates, these conditions imply that (conditional on x) blacks are sorted nonrandomly into positions of close friendship with whites.

If these conditions hold, then models that attempt to account for the observed association of interracial friendship and racial attitudes and also fail to control for nonrandom allocation of individuals into friendship statuses will yield inconsistent estimates of the effect of interracial friendship on racial attitudes.

MODEL 2: THE ASCRIPTION MODEL

The ascription model (Mare & Winship 1988) describes the situation in which persons are selected into statuses only on the basis of their observed characteristics. Under this model racial attitudes would not affect white friendship formation after controlling for covariates. If this prediction holds, then each equation estimated separately for the appropriate subsample would yield consistent estimates of covariate effects. This model is tested as a restricted version of the general model by setting the equation error covariances (correlations) to zero.

This model is consistent with the standard model of contact research; however, the effects of independent variables depend on whether the respondent reports having white friends.

MODEL 3: THE STRUCTURAL SHIFT MODEL OF HECKMAN

The two previous models allow covariate effects on racial attitudes to vary according to whether a person has white friends. This means not only that average sentiments are shifted by having close white friends, but that changes in the covariate values would exert nonproportional changes in sentiments when a person has close white friends. An alternative model proposed by Heckman (1978) allows only the *shift* in sentiments rather than the *change* in sentiments induced by differences in the effects of covariates. This model

implies that covariate effects are the same regardless of friendship status. The model is a restricted version of the general model obtained by constraining the regression slopes and error covariances (correlations) to be equal across the attitude equations, although equation intercepts are allowed to vary.

Heckman's model is consistent with the standard model of contact research (see, e.g., Ellison & Powers 1994; Sigelman & Welch 1993), but it allows for the possibility that interracial contact is an endogenous variable in a model predicting racial attitudes.

MODEL 4: THE STANDARD CONTACT MODEL

The exogenous contact model is the model invoked by most past research on the contact hypothesis. It can be cast as a restricted version of Heckman's model. The model is obtained by adding the constraint that equation covariances (correlations) are zero. That is, contact is assumed to be an exogenous determinant of racial attitudes.

In sum, the general model (model 1) and Heckman's structural shift model (model 3) run counter to the assumptions implicit in contact research. Both models incorporate parameters that permit an assessment of the magnitude and direction of selection bias. Model 3, however, constrains all covariate effects and error covariances or correlations to be equal regardless of whether a respondent has close white friends. Model 2 may be viewed as a weak version of model 4 (the standard contact model), obtained by including the full set of interactions of all independent variables with the close white friends variable. Stronger versions of model 2 would constitute a class of intermediate models incorporating a subset of interaction effects. The models may be summarized as follows:

		Interaction by Friendship Status	
		Yes	No
Sample Selection by Friendship Status	Yes	General (1)	Heckman (3)
	No	Ascription (2)	Standard (4)

MODEL SPECIFICATION

We impose various exclusion restrictions to assist in identifying the structural models and to ease the task of estimating the three-equation system. Specifically, we include certain variables in the selection equation (z) that are not included in the structural equations (y_0 and y_1). Substantively, these instrumental variables should tap general interpersonal attributes that facilitate friendship formation yet should be unrelated (or weakly related) to the formation of specific racial attitudes. Unlike most large national survey projects, the NSBA contains several interviewer assessments of the respondent and of aspects of the interview itself. For example, the interviewer was asked to rate the demeanor of the respondent at the time of the interview.[4] Two such interviewer assessments

are of interest in this study. Interviewers rated respondents on one continuum ranging from friendly (1) to hostile (7), and another continuum ranging from open (1) to suspicious (7). We use the average of these two items ($r = .72$) as an index of *interpersonal congeniality*. We include this post hoc interviewer evaluation as a predictor of close white friendships. This measure has the advantage of tapping the interpersonal qualities of the respondent and being a strong predictor of the formation of friendships with whites while not predicting racial attitudes. We also assume that gender affects friendship formation but has a negligible effect on skepticism. However, gender is assumed to influence attitudes toward interracial dating.

To summarize, the probability of forming friendships with whites is assumed to depend on age, urban residence, southern residence, education, income, childhood and adult interracial contact, interpersonal congeniality, and gender. Further, we assume that racial attitudes depend on all these factors except interpersonal congeniality and gender. Interracial dating attitudes are assumed to depend on all these factors except interpersonal congeniality.[5]

Results

Tables 1 and 2 present selected results from models of racial attitudes involving skepticism of whites' motives. Looking at the general model in Table 1, we find no evidence of nonrandom selection into white friendship status (for positive racial sentiments). The estimates of ρ_{1z} and ρ_{0z} are close to zero and statistically insignificant. Therefore, the basic exogenous contact model cannot be rejected on the basis of the data from the NSBA. Moreover, the ascription model with unconstrained slopes provides no evidence of the presence of interaction effects with friendship status. The findings in Table 2 for negative racial attitudes indicate that the exogenous selection model cannot be rejected, and the contact model fits as well as any of the less constrained models.

Table 3 presents selected results from the models of interracial dating attitudes. The estimates of σ_{1z} and σ_{0z} are close to zero and not statistically significant in both the general model and the structural shift model. Again, we find no evidence of selection bias. However, a likelihood chi-square test of the ascription model against the contact model shows that the former provides a better fit to the data ($\chi^2_9 = 37.4$).

Table 4 presents estimates from the general model of skepticism of the motives of whites. The effect of interracial contact (close white friends) is estimated as $\sum_j \hat{\beta}_{1j}\bar{x}_{1j} - \sum_j \hat{\beta}_{0j}\bar{x}_{0j}$. The effect of close white friends on positive racial attitudes is .686 while the effect on negative racial attitudes is -.338. Substantively, this implies that having white friends increases the probability of expressing optimism about the motivations of whites by .19 and decreases the chances of expressing pessimism about the motivations of whites by .13.[6]

Although the general model fits the data as well as the more parsimonious contact model, several differential effects of friendship status bear mentioning. First, the effect of income on *positive* attitudes among respondents with white friends is larger than the corresponding effect in the sample without white friends. On average, among those with white friends, a $1,500 increase in 1978

TABLE 1: Selected Results from Fitted Models: Dependent Variable Is the Positive Racial Attitude — "Whites Want to Give Blacks a Break"

	Model			
	General	Ascription	Heckman's Structural Shift	Contact
Selected estimates				
ρ_{1z}	0.002 (0.017)	0.0[c]	-0.006 (-0.077)	0.0[c]
ρ_{0z}	-0.014 (-0.099)	0.0[c]	-0.006[c]	0.0[c]
Log L	-4,203.019	-4,203.030	-4,206.987	-4,206.991
Degrees of freedom	31	29	22	21
χ^2	—	.022	7.963	7.944
Degrees of freedom	—	2	9	10

[a] Maximum likelihood estimates (asymptotic *t* statistics in parentheses)
[c] Denotes a constrained parameter

individual income leads to a .011 decrease in the probability of expressing the attitude that whites want to give blacks a break.[7] However, the *difference* in the income effect between persons with white friends and those without white friends is not statistically significant ($p = .38$), perhaps because of the small number of NSBA respondents without white friends who reported positive racial attitudes ($N = 89$). Second, income has a positive effect on the probability of expressing *negative* attitudes. Among respondents without close white friends, this effect is marginally significant ($p = .07$).[8] In the sample of NSBA respondents without white friends, a $1,500 increase in income leads to a .006 increase in the chances of expressing the attitude that whites want to keep blacks down. The magnitudes of these marginal effects suggest that income has, at best, a modest impact on skepticism attitudes.

Third, it is also noteworthy that among African Americans with white friends, age increases the likelihood of viewing the motivations of whites in a positive light. Each 10-year increase in age increases the chances of believing that whites want to give blacks a break by .039. The corresponding increase in probability for those without white friends is .030. This difference in age effects between persons with and without white friends is not statistically significant at conventional levels ($p = .11$). Our concluding section offers a more detailed substantive interpretation of these empirical patterns.

Table 5 shows how net effects would be estimated in the standard contact model. Overall, these effects are similar in magnitude to the estimates from the general model in Table 4 for NSBA respondents who reported having close

TABLE 2: Selected Results from Fitted Models: Dependent Variable Is Negative Racial Attitude — "Whites Want to Keep Blacks Down"

	Model			
Selected estimates	General	Ascription	Heckman's Structural Shift	Contact
ρ_{1z}	-0.017	0.0[c]	-0.006	0.0[c]
	(-0.184)	—	(-0.110)	
ρ_{0z}	0.002	0.0[c]	-0.006[c]	0.0[c]
	(0.031)	—	—	—
Log L	-4,861.889	-4,861.910	-4,863.129	-4,863.135
Degrees of freedom	31	29	22	21
χ^2	—	0.042	2.480	2.492
Degrees of freedom	—	2	9	10

[a] Maximum likelihood estimates (asymptotic t statistics in parentheses)
[c] Denotes a constrained parameter

white friends. The *effect* of close white friends in the contact model is estimated as the difference between the corresponding constant terms, or .649 for the positive racial attitude compared to -.341 for the negative racial attitude. (Alternatively, these effects could be estimated by including a term for close white friends in a single probit equation.) These effects imply that having white friends increases the probability of expressing optimism about the motivations of whites by .13 and decreases the probability of expressing pessimism about the motivations of whites by a comparable amount.

Table 6 compares estimates from the general model and the contact model of opposition toward interracial dating. The effect of close white friends on opposition to interracial dating is estimated as the difference in the average opposition to interracial dating between blacks who have close white friends and blacks who do not, or -.166. For the contact model, this effect is estimated by the difference between the constant terms, or -.148. The switching regressions approach permits a more complete assessment of covariate effects than the standard exogenous contact model. For example, the effect of education on opposition to interracial dating is three times as large among NSBA respondents with close white friends and is five times its standard error. These differences in the effects of education are statistically significant at conventional levels ($p = .001$).

We also find that casual adult interracial contact is associated with low levels of opposition to interracial dating among African Americans who have white friends. However, casual contact has no bearing on these attitudes among

TABLE 3: Selected Results from Fitted Models: Dependent Variable Is the Degree of Opposition toward Interracial Dating[a]

	Model			
	General	Ascription	Heckman's Structural Shift	Contact
Selected estimates				
σ_{1z}	-0.0097	0.0[c]	-0.0097	0.0[c]
	(-0.143)	—	(-.303)	
σ_{0z}	-0.0076	0.0[c]	-0.0097[c]	0.0[c]
	(-0.127)	—	—	—
Log L	-6,213.169	-6,213.203	-6,231.978	-6,232.020
Degrees of freedom	35	33	25	24
χ^2	—	0.068	37.618	37.702
Degrees of freedom	—	2	10	11

[a] Maximum likelihood estimates (asymptotic *t* statistics in parentheses)
[c] Denotes a constrained parameter

blacks without white friends. This difference between respondents with and without white friends is statistically significant at the .03 level. Thus, it seems that both the quantity and the variety of interracial contacts (intimate and casual) are important in eroding interracial suspicion and hostility.

The switching regressions approach offers some advantages over the conventional approach in revealing the differential effects of education and casual adult interracial contact on opposition to dating attitudes between those with and those without white friends. The net effect of allowing for differential effects can be seen by comparing the standard contact model to the ascription model. The likelihood ratio chi-square, or -2 times the difference between the log-likelihoods of the standard contact model and the ascription model in Table 3, is 37.4 with 9 degrees of freedom ($p < .005$). Therefore, the ascription model, or a less constrained version of the standard contact model, is more appropriate for explaining opposition toward interracial dating.

Summary and Conclusions

The contact hypothesis has been the subject of research in racial and ethnic relations for at least four decades. Although some recent investigators have found apparent support for the core claims of the contact hypothesis using data from large cross-sectional samples, they also raise the possibility that the empirical association between interracial contacts and positive racial attitudes

TABLE 4: Estimates from the General Model of Positive and Negative Racial Attitudes[a]

	Interracial Contact	Racial Attitude			
	White Friends	Give Break		Keep Down	
	$z = 1$	$y_{z=1} = 1$	$y_{z=0} = 1$	$y_{z=1} = 1$	$y_{z=0} = 1$
Constant	-1.752	-.871	-1.571	-.356	-.156
	(-9.10)	(-3.62)	(-3.92)	(-1.50)	(-.46)
Age/10	.137	.112	.090	-.007	.005
	(8.69)	(3.94)	(1.57)	(.26)	(.15)
Urban	-.111	-.112	-.041	-.113	-.104
	(-3.15)	(-.99)	-(.24)	(-1.42)	(-1.64)
South	.094	-.060	-.144	.072	.050
	(1.53)	(-.59)	(-.76)	(.08)	(.54)
Education/10	.284	.053	.011	.158	.115
	(1.86)	(.16)	(.03)	(.62)	(.45)
Income/10	.089	-.288	-.152	.096	.199
	(1.36)	(-2.58)	(-.70)	(.93)	(1.83)
Childhood interracial contact	.217	.010	.077	-.034	-.076
	(7.21)	(.41)	(.58)	(-.76)	(-.48)
Adult interracial contact	.134	.072	.031	-.006	.023
	(3.77)	(1.26)	(.29)	(-.12)	(.40)
Female	-.117	—	—	—	—
	(-2.33)				
Interpersonal congeniality	.103	—	—	—	—
	(5.04)				
ρ_{1z}, ρ_{0z}		.002	-.014	-.017	.002
		(.02)	(-.10)	(-.18)	(.03)
N	1,847	1,061	786	1,061	786
$N_{\text{dep var} = 1}$	1,061	317	89	386	389

[a] Maximum likelihood estimates (asymptotic t statistics in parentheses). The effect of missing data on income is estimated but is not reported.

TABLE 5: Estimates from the Contact Model of Positive and Negative Racial Attitudes[a]

	Interracial Contact	Racial Attitude			
	White Friends	Give Break		Keep Down	
	$z = 1$	$y_{z=1} = 1$	$y_{z=0} = 1$	$y_{z=1} = 1$	$y_{z=0} = 1$
Constant	-1.747 (-9.08)	-.858 (-4.67)	-1.507 (-8.40)	-.442 (-2.57)	-.101 (-.76)
Age/10	.137 (8.67)	.105 (5.72)	—	-.004 (.44)	—
Urban	-.112 (-3.15)	-.093 (-1.13)	—	-.108 (-2.14)	—
South	.094 (1.53)	-.095 (-.97)		.029 (.54)	
Education/10	.283 (1.85)	.041 (.32)	—	.147 (.82)	—
Income/10	.089 (1.35)	-.248 (-2.52)	—	.142 (1.88)	—
Childhood interracial contact	.216 (7.20)	.025 (.75)	—	-.046 (-1.43)	—
Adult interracial contact	.134 (3.72)	.055 (1.38)	—	.008 (.29)	—
Female	-.117 (-2.33)	—	—	—	—
Interpersonal congeniality	.103 (5.04)	—	—	—	—
ρ_{1z}, ρ_{0z}		—	—	—	—
N	1,847	1,061	786	1,061	786
$N_{\text{dep var} = 1}$	1,061	317	89	386	389

[a] Maximum likelihood estimates (asymptotic *t* statistics in parentheses). The effect of missing data on income is estimated but is not reported.

TABLE 6: Comparisons of Estimates from the General Model and Contact Model of Opposition towards Interracial Dating[a]

	Interracial Contact	Racial Attitude			
	White Friends	General Model		Contact Model	
	$z = 1$	$y_{z=1}$	$y_{z=0}$	$y_{z=1}$	$y_{z=0}$
Constant	-1.62 (-8.08)	2.869 (17.53)	2.760 (22.02)	2.780 (20.85)	2.928 (26.77)
Age/10	.139 (7.70)	.020 (1.43)	.010 (.61)	.017 (1.38)	—
Urban	-.147 (-3.20)	-.111 (-3.10)	-.174 (-2.39)	-.133 (-3.69)	—
South	.067 (1.33)	.067 (1.26)	-.091 (-1.44)	.016 (.42)	
Education/10	.431 (2.37)	-.811 (-5.80)	-.258 (-1.45)	-.599 (-4.66)	—
Income/10	.016 (1.08)	.042 (.62)	.065 (1.01)	.048 (1.14)	—
Childhood interracial contact	.204 (5.56)	-.040 (-1.66)	-.047 (-1.28)	-.049 (-2.46)	—
Adult interracial contact	.143 (3.39)	-.096 (-3.33)	-.021 (-.44)	-.074 (-3.09)	—
Female	-.143 (-2.49)	-.048 (-1.09)	.012 (.21)	-.026 (-.66)	—
Interpersonal congeniality	.089 (4.18)	—	—	—	—
σ_1, σ_0		.735 (41.15)	.816 (54.18)	.737 (41.17)	.823 (43.23)
$\sigma_{1z} \times 10$, $\sigma_{0z} \times 10$		-.097 (-.14)	-.076 (-.13)	—	—
N	1,720	988	732	988	732
$N_{\text{dep var}=1}$	988				

[a] Maximum likelihood estimates (asymptotic t statistics in parentheses). The effect of missing data on income is estimated but is not reported.

could result from unobserved selection processes. That is, preexisting positive racial attitudes may incline some individuals to form and sustain interracial friendships, while preexisting negative attitudes may discourage other persons from pursuing such interracial contacts. If the pivotal assumption that interracial contact is an exogenous determinant of racial attitudes is false, then future research on the contact hypothesis will be of little value.

One frequently mentioned strategy for examining this assumption of exogeneity would use longitudinal data (Jackman & Crane 1986; Sigelman & Welch 1993). However, a number of methodologists caution that longitudinal data may be unsuitable for resolving thorny issues of causality (Clogg 1986).[9] Moreover, to the best of our knowledge, such data are unavailable, especially for African Americans.

Therefore, we have taken a different tack, using endogenous switching regression models to analyze cross-sectional data from the National Survey of Black Americans (NSBA). Four competing models were elaborated and estimated: (1) the general selection model; (2) the ascription model; (3) Heckman's (1978) structural shift model; and (4) the standard contact model, which, like the ascription model, assumes that interracial contact is an exogenous determinant of racial attitudes. These analyses showed nothing that would lead us to reject the exogeneity assumption. There is no evidence that the observed association between interracial contact and racial attitudes is an artifact of unobserved selection processes, and hence there is no reason to reject the findings of previous research conducted within the contact tradition. It appears that while casual interracial contacts have little bearing on attitudes, close interracial friendships may provide African Americans with direct information about the values, life-styles, behaviors, and experiences of their white counterparts, enabling many to counter negative images and stereotypes of whites.

We emphasize that these methods may be generalized to a number of substantive areas of sociological inquiry in addition to the one considered here. Although we caution against the blind application of selection models in general, there are many situations in which methods to investigate selectivity bias are warranted. Researchers wanting to estimate treatment effects from nonexperimental data often face a missing data problem that makes causal inference subject to selection bias. The application of endogenous switching models to these situations would permit a more rigorous assessment of alternative assumptions about the processes that influence socioeconomic outcomes and sample selection probabilities.

Of course, this study has focused exclusively on African Americans and has examined only two of the many possible racial attitudes. Thus, further research is needed to test the exogeneity assumption for white Americans and for other attitudinal outcomes. Nevertheless, given these reassuring findings and the acrimonious contemporary racial climate, it is important to clarify those structural and institutional arrangements (e.g., in neighborhoods, schools, and workplaces) that promote the development of interracial friendships.

One ancillary benefit of our investigation of possible selection effects is that it reveals the contingent nature of the contact hypothesis. Research on the contact hypothesis has largely neglected selection effects, save for occasional attention to the role of various immediate situational factors — e.g., the views

of authorities on intergroup relations, the relative status of the parties engaging in contact, the cooperative or competitive nature of the contact — in moderating the relationship between contact and intergroup attitudes. Recent investigations using large secondary data sets have focused exclusively on estimating the main net effects of interracial contact on racial attitudes (e.g., Jackman & Crane 1986; Sigelman & Welch 1993), without considering the possibility that the nature and consequences might relate to one's age/cohort, socioeconomic status, or other aspects of social location.

Several intriguing patterns in our study suggest that this has been a nontrivial oversight. For instance, earlier research (Ellison & Powers 1994) reported an inverse association between educational attainment and black opposition to interracial dating, a finding that is consistent with most studies of white attitudes. However, our respecification of this relationship indicates that this relationship exists only among African Americans with close white friends; among their counterparts who lack these interracial bonds, there is no meaningful association between education and attitudes toward interracial dating (see Table 6). This difference in the estimated effects of education is large and statistically significant ($p < .001$).

This pattern is especially interesting because educational attainment is often considered a more accurate indicator of black middle-class standing than income (e.g., Smith 1989). However, a more modest pattern also emerges in the relationship between income and African American perceptions of whites' motives (Table 4). Among African Americans with close white friends, income reduces the likelihood that individuals will hold positive views of whites' motives but does not increase the likelihood that individuals will hold negative views. In contrast, among those persons lacking close interracial friendships, income is positively associated with the likelihood of believing that most whites "want to keep blacks down."

Although this latter difference does not reach conventional levels of statistical significance, perhaps because of the small number of NSBA respondents without white friends who reported positive perceptions of whites' motives, in both cases access to close interracial friendships seems to condition the link between socioeconomic status and African American racial attitudes. Our findings dovetail with several strands of recent research depicting the frustration and rage of many upwardly mobile blacks, who continue to combat various types of racial discrimination despite marked improvement in their economic resources and occupational standing. For instance, Feagin's (1991) interviews in several cities document the diverse covert and overt episodes of discrimination confronted in the course of daily life by well-educated, professional African Americans: rejection by lenders and realtors, poor service in restaurants and other business establishments, police harassment, and verbal epithets from white strangers, among many other examples. On the basis of interviews with leading executives, academics, and other upwardly mobile African Americans, Cose (1993) identifies twelve "demons," or recurring race-related stressors that have shattered expectations and reduced the quality of life for many middle- and upper-class African Americans. These items range from low expectations and lack of respect from white colleagues on the job to exclusion from white-dominated clubs and other social settings.

Taken together with evidence of "glass-ceiling" effects and other structural barriers to black middle-class success (e.g., Collins 1983, 1993; Geschwender & Carroll-Seguin 1990; Landry 1987), the research of Feagin, Cose, and others underscores the growing sense of frustration and relative deprivation among many upwardly mobile African Americans. Our findings suggest that the trust and openness of interracial friendships may be especially important in shaping the racial attitudes of members of the black middle class, reducing the likelihood that unpleasant encounters with interpersonal or institutional racism will be generalized into a negative portrait of whites as a group.

Two additional contingent patterns also merit brief discussion. First, among blacks with white friends, the casual contact that likely follows from involvement in racially integrated settings (e.g., workplaces, neighborhoods) appears to reduce opposition to interracial dating, an especially intimate form of interracial contact (Table 6). Among those blacks who lack close white friendships, on the other hand, we find no association between adult contact and attitudes toward interracial dating. This difference in the estimated effects of adult interracial contact is statistically significant ($p < .03$). It is interesting that this pattern does not emerge in analyses of perceptions of whites' motives in issues related to the welfare of African Americans (Table 4).

This finding raises the possibility that persons who acquire positive information about whites through close personal friendships may interpret subsequent exchanges and interactions with whites in a more positive light and may assimilate the information gained through these more casual contacts into a relatively benign image of whites as a group. However, due to the demographic imbalance between blacks and whites, most African Americans are forced to interact with whites from time to time regardless of their preferences for interracial contact (e.g., Butler & Wilson 1978). Thus, for persons who lack positive sources of information to counter widespread negative perceptions of whites, the level of casual contact has little bearing on attitudes toward interracial dating.

Our analyses have also turned up limited evidence of age/cohort differences in the implications of interracial friendships for African American attitudes. Among blacks with white friends, older respondents tend to report favorable views of whites' motives (Table 4), while among blacks who lack close interracial friendships, age is unrelated to this attitudinal outcome. There are two plausible interpretations of this pattern. On the one hand, by virtue of their age, older African Americans have had longer to cultivate and nurture close friendships with whites. It is reasonable to suspect that interracial friendships of longer duration may provide more candid and revealing information about whites than those of more recent origin. On the other hand, this pattern may constitute modest evidence of a cohort effect. Older African Americans may have vivid recollections of earlier periods in which race relations were more acrimonious and racial barriers were more rigid, and they may also recall the significant interracial cooperation of the civil-rights era. Given this historical frame of reference, older African Americans may more readily generalize positive experiences with whites gained via personal friendships into a relatively favorable image of whites as a group. Possible age/cohort differences in the effects of interracial contact merit further research.

This study has augmented our understanding of the causes, processes, and consequences of interracial contact by (1) reaffirming the key causal assumption underlying the contact hypothesis, (2) providing new information about the relationship between interracial contact and the racial attitudes of African Americans, and (3) demonstrating that the effects of such contact vary according to aspects of social location (e.g., socioeconomic status, age/cohort). Additional research is needed to explore these contingent relationships in a broader array of racial attitudes and to investigate the existence of such contingent patterns among white Americans as well. Finally, the NSBA data were collected from 1979 to 1980, and substantial changes in the U.S. political economy and racial climate over the last 15 years may have altered the patterns identified in this study. Although, to the best of our knowledge, more recent data on interracial contact and racial attitudes in large samples of African Americans are exceedingly rare, the collection and analysis of such data should be an important priority for the future.

APPENDIX

Each competing model imposes a different set of restrictions on equation slope parameters β_{0j}, β_{1j}, γ_j, $(j = 1,k)$ η_0 and η_1, or equivalently, equation covariances σ_{1z} and σ_{0z} or correlations ρ_{1z} and ρ_{0z}.

THE GENERAL MODEL

The general model imposes the fewest restrictions on parameters. The only assumption is that σ_{10} or $\rho_{10} = 0$, whereas σ_{1z}, σ_{0z}, ρ_{1z}, and ρ_{0z} are free to vary.

THE ASCRIPTION MODEL

The ascription model is tested as a restricted version of the general model by setting $\sigma_{1z} = \sigma_{0z} = 0$ or $\rho_{1z} = \rho_{0z} = 0$.

THE STRUCTURAL SHIFT MODEL OF HECKMAN

Heckman's structural shift model is a restricted version of the general model obtained by setting $\beta_{0j} = \beta_{1j}$ $(j > 1)$ and $\sigma_{1z} = \sigma_{0z}$ or $\rho_{1z} = \rho_{0z}$.

THE STANDARD CONTACT MODEL

The exogenous selection contact model can be cast as a restricted version of Heckman's model. In addition to the constraint that the $\beta_{0j} = \beta_{1j}$ $(j > 1)$, this model is obtained by setting $\eta_0 = \eta_1 = 0$, or equivalently, $\sigma_{1z} = \rho_{1z} = \sigma_{0z} = \rho_{0z} = 0$.

All these models can be estimated in HotzTran (Avery & Hotz 1985). The ascription model and standard contact model can be estimated using OLS and probit regressions when the observed attitudinal outcome is continuous or binary. The structural shift model can be estimated using bivariate probit regressions.

Notes

1. Mare and Winship (1988) discuss the situation in which assignment to position is made by a decision maker (e.g., the placement of a student in a particular school track level is made by teachers, school administrators, and parents). Our situation is one in which individuals self-select into different positions.

2. This is a latent variable approach insofar as y_1 and y_0 may be thought of as blacks' propensity to exhibit positive (or negative) perceptions of whites' motivations, whereas z may be thought of as blacks' propensity to form friendships with whites.

3. Identification of this model hinges on the assumption of a trivariate normal error distribution. The realizations of the latent dependent variables in equations 1a and 1b may be continuous, binary, or censored outcomes, corresponding to OLS, probit or tobit equations respectively (Powers 1993).

4. All NSBA interviews were conducted by African Americans.

5. A number of alternative models were examined that included different exclusion restrictions or no exclusion restrictions. The estimates we report are robust to alternative model specifications.

6. The marginal effect of close white friendship on the probability of expressing a particular racial attitude is given by $\Phi(\sum_j \hat{\beta}_{1j}\bar{x}_{1j}) - \Phi(\sum_j \hat{\beta}_{0j}\bar{x}_{0j})$, where $\Phi(.)$ denotes the standard cumulative normal distribution function (cdf).

7. The marginal effect of a change in x_k on the probability that $y = 1$ is given by $\partial P(y = 1|x_k)/\partial x_k = \phi(\sum x_k\hat{\beta}_k)\hat{\beta}_k$, where $\phi(.)$ denotes the standard normal probability density function (pdf). Individual income is constructed as a 17-point summary scale. A unit change in the scale represents an average change of $1,500 in 1978 dollars.

8. All reported p values are for two-tailed tests.

9. Although many authors concur that the relationship between contact and attitudes is bidirectional, most assume that contact causes attitudes, not the reverse. Whether causality can be assessed in the absence of longitudinal data is questionable. The question of the exogeneity of the effect of interracial contact would remain an issue regardless of the sampling plan.

References

Allport, Gordon W. 1954. *The Nature of Prejudice*. Beacon Press.

Avery, Robert B., and V. Joseph Hotz. 1985. *HotzTran User's Manual*. CERA Economic Consultants.

Blalock, Hubert M., Jr. 1964. *Causal Inferences in Nonexperimental Research*. University of North Carolina Press.

Brown, Barry, and George Albee. 1966. "The Effect of Integrated Hospital Experiences on Racial Attitudes." *Social Problems* 13:324-33.

Butler, John Sibley, and Kenneth L. Wilson. 1978. "The American Soldier Revisited: Race Relations and the Military." *Social Science Quarterly* 59:451-67.

Clogg, Clifford C. 1986. "Invoked by RATE." *American Journal of Sociology* 92:696-706.

Collins, Sharon. 1983. "The Making of the Black Middle Class." *Social Problems* 30:369-82.

____. 1993. "Blacks on the Bubble: The Vulnerability of Black Executives in White Corporations." *Sociological Quarterly* 34:429-48.

Cose, Ellis. 1993. *The Rage of a Privileged Class*. Harper & Row.

Demo, David H., and Michael Hughes. 1990. "Socialization and Racial Identity among Black Americans." *Social Psychology Quarterly* 53:364-74.

Desforges, Donna H., Charles G. Lord, Shawna L. Ramsey, Julie A. Mason, Marilyn D. Van Leeuwen, Sylvia C. West, and Mark P. Lepper. 1991. "Effects of Structured Cooperative Contact on Changing Negative Attitudes toward Stigmatized Groups." *Journal of Personality and Social Psychology* 60:531-44.

Ellison, Christopher G. 1991. "Identification and Separatism: Religious Involvement and Racial Orientations among Black Americans." *Sociological Quarterly* 32:477-94.

Ellison, Christopher G., and Daniel A. Powers. 1994. "The Contact Hypothesis and Racial Attitudes among Black Americans." *Social Science Quarterly* 75:385-400.

Feagin, Joe. 1991. "The Continuing Significance of Race: Antiblack Discrimination in Public Places." *American Sociological Review* 56:101-16.

Ford, W. Scott. 1973. "Interracial Public Housing in a Border City." *American Journal of Sociology* 78:426-44.

Geschwender, James, and Rita Carroll-Seguin. 1990. "Exploding the Myth of African-American Progress." *Signs* 15:285-99.

Heckman, James, J. 1978. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica* 46:931-59.

Jackman, Mary R., and Marie Crane. 1986. "'Some of My Best Friends Are Black': Interracial Friendship and Whites' Racial Attitudes." *Public Opinion Quarterly* 50:459-86.

Jackson, James S. 1991. "Methodological Approaches." Pp. 13-30 in *Life in Black America,* edited by James S. Jackson. Sage.

Jackson, James S., Belinda Tucker, and Gerald Gurin. 1987. National Survey of Black Americans, 1979-1980 [MRDF]. Ann Arbor: University of Michigan, Survey Research Center.

Landry, Bart M. 1987. *The New Black Middle Class.* University of California Press.

Mare, Robert D., and Christopher Winship. 1988. "Endogenous Switching Regression Models for the Causes and Effects of Discrete Variables." Pp. 132-60 in *Common Problems/Proper Solutions: Avoiding Error in Quantitative Research,* edited by J. Scott Long. Sage.

McDowell, Sophia. 1971. "Patterns of Preference by Negro Youth for White and Negro Associates." *Phylon* 32:290-301.

Powers, Daniel A. 1993. "Endogenous Switching Regression Models with Limited Dependent Variables." *Sociological Methods and Research* 22:248-73.

Robinson, Jerry W., and James D. Preston. 1976. "Equal-Status Contact and Modification of Racial Prejudice: A Reexamination of the Contact Hypothesis." *Social Forces* 54:911-24.

Schuman, Howard, and Shirley Hatchett. 1974. *Black Racial Attitudes.* University of Michigan, Institute for Social Research.

Schuman, Howard, Charlotte Steeh, and Lawrence Bobo. 1988. *Racial Attitudes in America: Trends and Interpretations.* 2d ed. Harvard University Press.

Sigelman Lee, and Susan Welch. 1993. "The Contact Hypothesis Revisited: Interracial Contact and Positive Racial Attitudes." *Social Forces* 71:781-95.

Smith, A. Wade. 1989. "Educational Attainment as a Determinant of Social Class among Black Americans." *Journal of Negro Education* 58:416-29.

Stephan, Walter G. 1987. "The Contact Hypothesis in Intergroup Relations." Pp. 13-40 in *Group Processes and Intergroup Relations,* edited by Clyde Hendrick. Sage.

Wilner, Daniel M., Rosabelle P. Walkley, and Stuart W. Cook. 1955. *Human Relations in Interracial Housing: A Study of the Contact Hypothesis.* University of Minnesota Press.

Winship, Christopher, and Robert D. Mare. 1992. "Models for Sample Selection Bias." *Annual Review of Sociology* 18:327-50.

Works, Ernest. 1961. "The Prejudice Interaction Hypothesis from the Point of View of the Negro Minority Group." *American Journal of Sociology* 67:47-52.

Exploring the Determinants of Adult Black Identity: Context and Process*

DAVID HARRIS, *Northwestern University*

Abstract

This article assesses the impact of childhood interracial contact and parental socialization on adult black identity. The investigation goes beyond previous work by testing hypotheses about the process by which these two factors operate, as well as differentiating between black and person-of-color identities. Evidence is found to support claims that interracial contact in childhood weakens adult feelings of closeness to other blacks. Mixed support for the impact of parental socialization emerges once a more precise definition of identity is employed. With respect to process, insufficient support is found for the claim that childhood factors affect closeness through black culture and outgroup orientation. The article concludes by outlining arguments regarding the merits of black identity and calling for research to help resolve this controversy.

This article addresses concerns that accompany integration by examining the impact of childhood interracial contact and socialization messages on adult black identity.[1] While this topic has previously been the subject of several studies (e.g., Broman, Jackson & Neighbors 1989; Demo & Hughes 1990; Jackson, McCullough & Gurin 1988), the present analysis represents an extension of this research in two important ways. First, the article contributes to the literature on black identity by introducing a refinement of the concept of identity. The closeness measure used in other work is modified to differentiate black identity from feelings of closeness to other people of color. Second, both direct and indirect effects of childhood factors on adult black identity are examined. Too often previous work has identified important determinants of identity but failed to empirically address process. However, if we are to expand our knowledge of racial identity, the importance of understanding process is certainly on par with that of assessing the significance of potential determinants.

* *The data used in this article were made available by the Inter-University Consortium for Political and Social Research. The data for the National Survey of Black Americans, 1979-1980, were originally collected by James S. Jackson and Gerald Gurin. This research was supported by funding from the National Science Foundation. I wish to thank Roberto Fernandez, Nicola Beisel, Aldon Morris, Christopher Jencks, and Michael Thornton for their helpful comments on earlier drafts of this article. Each helped greatly to improve the work presented here, but are of course not responsible for its content. Direct correspondence to David Harris, Department of Sociology, Northwestern University, Evanston, IL 60208-1330.*

Data and Methods

THE SAMPLE

Data for this project are drawn from the National Survey of Black Americans (NSBA; Jackson & Gurin 1987). Collected by researchers at the University of Michigan, this data set is the result of interviews conducted between 1979 and 1980 with 2,107 respondents. The age range of respondents is from 17 to 101. This national household probability survey was administered face to face by black interviewers. For the most part, the NSBA is representative of the national black population enumerated in the 1980 census. Exceptions are a slight overrepresentation of females and older blacks and a small underrepresentation of southerners.

Variables

DEPENDENT VARIABLE

In this article, black identity is represented by feelings of closeness to other blacks. This conceptualization of identity conflicts with two findings from the literature. First, black identity is a multidimensional construct. Demo and Hughes (1990) and Phinney (1990) report feelings of closeness, separatist notions, and group evaluation as distinct components of black identity. The present concentration on only one of these dimensions is not meant as a dismissal of the other two, but rather is an attempt to emphasize the facet of black identity that most closely resembles classic definitions of group identity. The closeness measure employed here captures the crucial characteristics of group identity as "we" (Cooley 1902) and "a shared feeling of peoplehood" (Gordon 1964: 24) better than do constructs of separatism or group evaluation.

Second, several scholars have shown the strength of black identity to depend on the segment of the black population chosen as the reference group (e.g., Allen, Dawson & Brown 1989; DuBois [1899] 1973). While such attention to the intricacies of intraracial identity are certainly important, there is also a great need to understand blacks' identification with all other blacks. Many issues affecting blacks are not specific to class, age, or gender. Most people who see a black person on the street are unable to surmise much about the individual other than his or her race (Feagin 1991). As many scholars have noted, blackness is like a uniform. Especially in the company of strangers, it has the tendency to supersede an individual's other characteristics (Blauner 1972; Gordon 1964; Park 1950). For this reason the present analysis does not divide black identity into feelings of closeness to various intraracial reference groups.

Operationalizing black identity as closeness to other blacks is nothing new. In fact, it is a definition selected by many scholars who have worked with the NSBA. Where the construction of identity in this article differs from previous work is in its effort to purge black identity of its component that represents a sense of community with other nonwhites. Blauner (1972) maintains that nonwhite Americans think of themselves as a group in opposition to whites, rather than as an array of distinct ethnicities. He attributes this conception of

society to the civil-rights struggles of the 1960s. For many blacks, Hispanics, Native Americans, and Asian Americans, the lesson has been learned that minority groups must work together if they are ever to secure a position of equality with whites. This merger of interests has been accompanied by the appearance of a new group in society — people of color. It is for this reason that all equations in which closeness to blacks is the dependent variable also include a control for closeness to other minorities. Without such a control, it is unclear whether observed effects are on one's black or person-of-color identities.

Feelings of Closeness to Other Blacks

Respondents were asked how close they felt in their ideas and feelings to blacks who are (1) poor, (2) religious, (3) young, (4) middle-class, (5) working-class, (6) older, (7) elected officials, and (8) professionals. Possible responses were (1) not at all close, (2) not too close, (3) fairly close, and (4) very close. The raw measure of closeness to other blacks is the average response to the eight questions above, with a high value corresponding to greater closeness. If a respondent had a missing value for three or fewer categories, the mean response was used for those dimensions. Those with more than three missing responses were excluded from analyses. The alpha reliability coefficient for this scale is .81 (Broman, Neighbors & Jackson 1988).

Feelings of Closeness to Other Minorities

In addition to questions about closeness to blacks, respondents were asked how close they feel to (1) Spanish-speaking groups in the U.S., (2) American Indians, and (3) Asian Americans. Acceptable responses were (1) not at all close, (2) not too close, (3) fairly close, and (4) very close. The summary measure of closeness to other U.S. minority groups is the average answer to these three questions. Respondents with more than one missing value on the three questions were not included in further analyses.

Independent Variables

Childhood interracial contact is an attempt to quantify the level of contact an individual experiences with white people from birth to the beginning of adulthood. The NSBA asked respondents to estimate the racial composition of their (1) grammar school, (2) junior high, (3) high school, (4) college, and (5) neighborhood where they lived when growing up. Acceptable responses were (1) all black, (2) mostly black, (3) about half black, (4) mostly white, and (5) almost all white. The value of this variable is obtained by averaging the total score from the five questions across the number of categories with valid responses. Only those with missing data for all domains were dropped from analyses (Demo & Hughes 1990; Hughes & Demo 1989).

Parental socialization:[2] Respondents with children were asked if they have told their children anything about what it means to be black or how to get along with whites. If so, the content of those messages was obtained. Next, all respondents were asked if their parents had told them anything about what it means to be black or how to get along with whites when they were children. Those who said "no" were coded as having received no socialization. Respon-

dents who said "yes" and acknowledged socializing their children in the same way received a score corresponding to the message communicated to their children. All others were questioned about the content of messages received from their parents and those responses were coded.

The first step in converting these responses to a childhood socialization variable was to divide responses into three categories. The first group is labeled individualistic/universalistic. These respondents were given advice that denied the importance of race. The second is called integrative/assertive. These people's parents told them to maintain a positive group-oriented outlook. The third category is cautious/defensive. This group received messages about white prejudice and the importance of black power. The fourth group did not receive any socialization messages.

Next, three dummy variables were created. Each corresponds to one of the three types of socialization messages specified above and is coded 1 if the respondent received that type of message and 0 otherwise. The final step was to add the dummy variables that refer to the same type of message but to tally separately messages that made reference to blacks and those that made reference to whites. The result of this process is the creation of three measures of childhood socialization: individualistic/universalistic, integrative/assertive, and cautious/defensive. Values for each are (0) no messages of that type, (1) messages of that type about either what it means to be black or how to get along with whites, and (2) messages of that type about what it means to be black and how to get along with whites.

Intervening Variables

Outgroup orientation is represented by two measures and is based on the work of Jackson et al. (1991). The first is a global assessment of how whites view blacks. Interviewers asked, "On the whole, do you think most white people want to see blacks get a better break, or do they want to keep blacks down, or don't they care one way or the other?" Those who believed whites want to see blacks get a better break are coded 1. Respondents who answered either that "whites want to keep blacks down" or that "whites don't care either way" are coded 0.

The second measure of outgroup orientation considers respondents' personal networks. Interviewers asked, "Do you know any white person who you think of as a good friend — that is, someone to whom you can say what you really think?" Responses of "yes" are represented by a 1, and those of "no" are classified as 0.[3]

Traditional black church:[4] Respondents were asked whether they currently attend church and if so, how often. From the pool of people who reported attending church services at least once a month, individuals belonging to traditional black denominations (major national Baptist organizations, United Methodist, Christian Methodist Episcopal, African Methodist Episcopal, and African Methodist Episcopal Zion) are coded as 1 on this variable. Respondents who either attend church infrequently or are members of denominations other than those five are coded 0 (Ellison 1991).

Control Variables

Parents' socioeconomic status (SES): In the NSBA, respondents were asked the primary occupation of each parent when they were a child. A second question ascertained the parent's level of schooling. Due to the fact that so many respondents did not answer the education question, this information cannot be incorporated into a measure of SES. A further problem in determining socioeconomic status is that the NSBA provides neither a detailed occupational classification for respondents' parents nor enough information to construct one. The only information for assigning SES is a very broad grouping of parental jobs into one of seven categories: (0) no occupation; (1) farmer; (2) service worker or unskilled laborer; (3) operative, including transport; (4) craftsman, foreman, government protective service worker, or member of the armed forces; (5) clerical or sales worker; and (6) professional.[5] In the case that both parents were employed when the respondent was a child, only the occupation with the highest code is used (Demo & Hughes 1990; Hughes & Hertel 1990).

Complexion: This variable represents the direct transformation of a question interviewers were asked to answer. They recorded the respondent's skin color as (1) very dark brown, (2) dark brown, (3) medium brown, (4) light brown, or (5) very light brown.

Age: This is the respondent's reported age in years.

Female: This variable has a value of 0 for males and 1 for females.

South: Respondents were asked to name the state where they lived during most of their childhood. This information was converted into a dummy variable with 1 representing childhood residence in the South.

Results

The mean score on reported feelings of closeness to other blacks is 3.39 (maximum closeness = 4). This translates to a response between "fairly" and "very" close to blacks. In fact, only 1% of blacks reported feelings below 2 ("not too close"). The average respondent did not feel as close to other minorities as he or she did to other blacks. The mean closeness to other minorities is 2.11. The fact that the vast majority of respondents expressed greater solidarity with other blacks than with nonblack minorities suggests that what is examined in the course of this article is not the presence or absence of black identity, but variations in its degree.

In order to evaluate the total effects of childhood interracial contact and parental socialization on adult feelings of closeness to other blacks, ordinary least-squares regression coefficients for the reduced form equation are presented in model 1 of Table 1.[6] As predicted, the relationship between closeness to other blacks and interracial contact is negative and strongly significant. The coefficient for contact predicts that a change in childhood environment from "all black" to "almost all white" would result in the weakening of feelings of closeness by approximately one-sixth of one point. While this reduction is statistically significant, it may not be substantively important. We need to know the consequences for the individual and for the black community of changes in

feelings of closeness. This is a question that has not been adequately addressed in the literature to date, but it will be explored in the final section of this article.

To demonstrate the effect of controlling for closeness to other minorities on the apparent influence of childhood factors, model 2 is presented, which conforms to previous constructions of black identity (e.g., Broman, Neighbors & Jackson 1988; Demo & Hughes 1990) by not including person-of-color identity. Interracial contact has a greater negative effect on the isolated measure of black identity in model 1 than in model 2. All three measures of parental socialization have a stronger effect in model 2, thus substantiating the claim that messages about blacks and whites partially influence one's person-of-color identity.

Model 1 also indicates that parental socialization exerts an inconsistent influence on adult feelings of closeness to other blacks, a finding that contradicts past work (Demo & Hughes 1990; Jackson et al. 1991). While these studies conclude that all varieties of parental socialization influence closeness, the present analysis finds evidence of a positive effect of integrative/assertive and individualistic/universalistic parental messages, but no impact from cautious/defensive teachings about race. This discrepancy is the result of a difference in the construction of the closeness measure. The fact that messages exhorting a defensive posture toward whites are shown here to be inconsequential for black identity suggests their effect is on one's feelings of closeness to other people of color. Thus, when feelings of closeness to other minorities are controlled, as in model 1, the effect of this factor disappears.

Finally, it is of interest to note that model 1 has an R^2 of .2057. The fact that 21% of the variance in adult closeness to other blacks can be explained with childhood variables, closeness to other minorities, and demographic controls points to the importance of early experiences. However, it is also true that 79% of the variance in closeness is unaccounted for by the current model. While some of this unexplained variance is likely due to data noise and measurement error, it is improbable that these factors account for all the remaining variance. Therefore, it is reasonable to conclude that the effects observed for interracial contact and parental messages might be overwhelmed by other factors (e.g., adult characteristics or experiences) not included in this analysis.

Now that the magnitude and direction of the effects of childhood interracial contact and parental socialization on adult feelings of closeness to other blacks have been assessed, the analysis shifts to address the article's other stated goal — an evaluation of the process by which childhood factors exert their influence. The hypothesis is that interracial contact and parental socialization affect closeness indirectly, through outgroup orientation and participation in black churches, as well as directly.

The first step in evaluating this model is to examine the effect of childhood factors on black church involvement and outgroup orientation. Table 2 reports these results. As anticipated, increased interracial contact serves to lessen one's involvement with black churches. The move from an all-black to mostly black environment results in only a 9% decrease in the likelihood of attending a black church. However, the effect is not significant. The lack of a strong effect of interracial contact challenges the claim that involvement in a group's cultural activities is dependent on the extent to which that group exists in the individual's ecological range. As blacks move to more integrated environments,

TABLE 1: OLS Regression Coefficients for Reduced Form Model of Closeness to Other Blacks

	Model 1		Model 2	
	b	Beta	b	Beta
Interracial contact	-.0419** (.0160)	-.0763	-.0353* (.0165)	-.0641
Parental socialization				
Individualistic/ universalistic	.0522* (.0207)	.0629	.0730** (.0212)	.0881
Integrative/ assertive	.0958** (.0197)	.1201	.1088** (.0203)	.1364
Cautious/ defensive	.0394 (.0210)	.0473	.0550* (.0216)	.0660
Controls				
Parents' SES	-.0155 (.0080)	-.0485	-.0145 (.0082)	-.0452
Complexion	-.0257* (.0128)	- .0477	-.0252 (.0132)	- .0466
Female	-.0010 (.0243)	-.0010	-.0372 (.0248)	-.0365
Age	.0048** (.0007)	.1710	.0055** (.0007)	.1971
South	.2081** (.0304)	.1948	.1953** (.0314)	.1828
Closeness to other minorities	.1495** (.0155)	.2327	— —	—
Constant	2.8342** (.0788)		3.1142** (.0756)	
Adjusted R^2	.2057		.1548	
N	1,452		1,452	

[a] Numbers in parentheses are standard errors.

* $p < .05$ ** $p < .01$

theory predicts that black churches become less relevant for negotiating the matters of daily life and are replaced by class-based cultural manifestations (Gordon 1964; Stinchcombe 1975). No evidence is present in this analysis to suggest the abandonment of black institutions by integrated blacks.

Table 2 also indicates a positive effect of parental socialization on involvement with a traditionally black church. In fact, all three types of messages serve to increase the likelihood of church activity, though none of the estimates are statistically reliable at the .05 level. Lessons extolling the merits of blacks and encouraging interracial cooperation have the greatest effect, leading to a 20% increase in the likelihood of participation. The least important class of socialization messages, those that are cautious or defensive, has an effect about half as strong.[7]

The second and third columns of Table 2 present information about the effect of childhood factors on outgroup orientation. Beliefs about whether whites want to see blacks get better opportunities reflect a global assessment, while the presence of a white friend illustrates personal outgroup orientation. In support of the hypothesis that contact inspires comfort with one's outgroup (Gordon 1964; Park 1950), the coefficients for interracial contact are positive and significant for both global and personal assessments of whites. Despite the agreement in direction, there is a difference in the strength of these relationships. The change from complete segregation to a mostly black environment results in a 61% increase in the likelihood of reporting personal comfort, but only a 23% rise in the likelihood of expressing a positive global assessment of whites. The large increase in the likelihood of reporting a white friend is to be expected, since increased integration leads to a greater number of potential white friends. At school, the local playground, and neighborhood clubs, integrated black children interact and form primary relationships with many white boys and girls (Gordon 1964). The fact that such individuals continue to participate in relationships with whites into adulthood attests to the lasting impact of early positive interracial experiences.

Unlike interracial contact, the coefficients in Table 2 indicate mixed agreement between the predicted and observed effects of parental socialization on outgroup orientation. Individualistic messages increase the likelihood of a positive personal outgroup orientation by 28% but have little influence on global assessments. The latter finding is understandable in that this type of socialization is decidedly lacking in either positive or negative racial sentiments. While such messages do not affect ideology, they do matter for day-to-day decisions of who will be chosen as a friend. Individualistic renderings promote evaluations solely on the basis of character, thereby encouraging relationships with people of all races.

The second type of socialization message, one that is integrative in nature, has the strongest effect on both orientation measures. It increases both by approximately 40%. This finding is in agreement with previous work (Jackson, McCullough & Gurin 1988; Jackson et al. 1991), and indicates the long-term effectiveness of a philosophy promoting interaction between whites and blacks.

Finally, messages of a defensive and cautious nature increase the likelihood of a positive personal outgroup orientation by 33% but have an opposite though statistically insignificant effect on global comfort. This discrepancy suggests

TABLE 2: Logistic Regression Coefficients for Intervening Variables

	Traditional Black Church		Outgroup Orientation: Whites Support Blacks		Outgroup Orientation: White Friend	
	b	exp(b)	b	exp(b)	b	exp(b)
Interracial contact	-.0935	.9108	.2059*	1.2286	.4742**	1.6067
	(.0770)		(.0888)		(.0811)	
Parental socialization						
Individualistic/ universalistic	.1684	1.1834	.0563	1.0579	.2461*	1.2790
	(.0988)		(.1139)		(.0984)	
Integrative/assertive	.1794	1.1965	.3618**	1.4359	.3164**	1.3722
	(.0945)		(.1050)		(.0948)	
Cautious/defensive	.0928	1.0973	- .2027	.8165	.2889**	1.3349
	(.1002)		(.1252)		(.0999)	
Controls						
Parents' SES	-.0259	.9744	.0307	1.0311	.0781*	1.0812
	(.0382)		(.0448)		(.0383)	
Complexion	-.0866	.9170	-.1050	.9003	.0393	1.0401
	(.0618)		(.0716)		(.0612)	
Female	.5004**	1.6493	.1001	1.1053	-.2126	.8085
	(.1156)		(.1356)		(.1149)	
Age	.0146**	1.0147	.0224**	1.0226	.0179**	1.0181
	(.0034)		(.0039)		(.0035)	
South	.8797**	2.4102	.1916	1.2112	.2543	1.2896
	(.1461)		(.1779)		(.1446)	
Constant	-1.2144**		-2.7543**		-1.9524**	
	(.3550)		(.4225)		(.3587)	
χ^2	145.407		56.197		82.130	
N	1,452		1,452		1,452	

[a] Numbers in parentheses are standard errors.

* $p < .05$ ** $p < .01$

parents who preach to their children about the omnipresent threat of racism have their greatest influence at the level of ideology. Instilling caution about whites on the level of personal interaction actually serves to foster a positive outgroup orientation. The reason for this association is not clear but may be a result of the mismatch between the degree of racism predicted by parents and that actually experienced by the child.

Table 3 presents the results of an OLS regression equation evaluating the full model suggested earlier in this article. Involvement with a traditionally black church is significantly related to black identity in the direction predicted by previous work (Ellison 1991). Blacks who regularly attend such institutions report feelings of closeness that are one-eighth of a level stronger than the sentiments expressed by fellow ethnics who are not active members of black churches. The fact that this relationship holds even with the constrained operationalization of identity employed in this analysis indicates that the black church is effective in encouraging a distinctly black identity.

Table 3 also provides information about the impact of outgroup orientation on adult feelings of closeness to other blacks. Contrary to expectations, there is no evidence that outgroup orientation negatively affects ingroup closeness. Neither dimension shows a statistically reliable estimate, and only the effect of global orientation is in the hypothesized direction. This finding conflicts with views expressed by Abrams and Hogg (1990) and Cooley (1902) that feelings of closeness are strongest when there are positive ingroup and negative outgroup sentiments. The absence of predicted effects of outgroup orientation suggests that these theories of group identity lack relevance to African Americans and indicates that the strength of black identity is independent of feelings about whites.

As predicted, the inclusion of black church participation and outgroup orientation in the model of adult feelings of closeness to other blacks reduces the observed effects of interracial contact and parental socialization. However, the magnitude of these changes is very small. The coefficient on childhood interracial contact is reduced by only 0.48%. The three measures of parental socialization fall by an average of 10%. That the indirect effects of these variables are so small testifies to the fact that outgroup orientation and participation in a traditional black church are not the mechanisms by which the examined childhood factors affect adult feelings of closeness to other blacks. Clearly, other processes are at work.

Discussion and Conclusions

This study has sought to answer two important questions about black identity. First, what are the effects of childhood racial environment and parental socialization on adult black identity? Past research has been clear in its conclusion that both of these factors are significant determinants of black identity. The study provides ample evidence for the claim that interracial contact weakens racial identity but finds mixed support for the argument that parental socialization strongly affects feelings of closeness to other blacks.

TABLE 3: OLS Regression Coefficients for Full Model of Closeness to Other Blacks

	b	Beta
Interracial contact	-.0417** (.0161)	-.0758
Parental socialization		
Individualistic/universalistic	.0457* (.0205)	.0551
Integrative/assertive	.0897** (.0197)	.1124
Cautious/defensive	.0341 (.0209)	.0409
Intervening variables		
Traditional black church	.1213** (.0241)	.1233
Outgroup orientation		
Whites support blacks	-.0160 (.0288)	-.0135
White friend	.0273 (.0246)	.0274
Controls		
Parents' SES	-.0152 (.0079)	-.0476
Complexion	-.0239 (.0127)	-.0443
Female	-.0127 (.0243)	-.0124
Age	.0043** (.0007)	.1543
South	.1821** (.0306)	.1705
Closeness to other minorities	.1520** (.0155)	.2365
Constant	2.8009** (.0785)	
Adjusted R^2	.2187	
N	1,452	

[a] Standard errors are in parentheses.

* $p < .05$ ** $p < .01$

An explanation for this discrepancy was found in the definitions of black identity employed by different studies. Although defining identity in terms of feelings of closeness was seen to be quite common, the current study distinguished black identity per se from identification with other people of color. The importance of this distinction is that it allows factors affecting black identity to be distinguished from those relevant to one's nonwhite identity. Thus, the lack of consensus between this and other studies can be seen as an indication that warnings about omnipresent racial discrimination are important only for one's identity as a person of color. Previous findings that this class of parental messages is a determinant of feelings of closeness to other blacks are the result of poor specifications of black identity.

The conclusion that interracial contact, and to a lesser degree parental socialization, affect adult feelings of closeness to other blacks leads to an assessment of the second question motivating this research — what is the process by which childhood factors influence adult identity? Theory suggests that group identity is the result of positive ingroup sentiments and a negative outgroup orientation.

The present study provides insufficient evidence to support the conclusion that ingroup culture and outgroup orientation are the mechanisms by which childhood factors affect adult black identity. Consideration of indirect effects through these two mechanisms resulted in a minimal reduction in the strength of the direct effects of interracial contact and parental socialization. Furthermore, outgroup orientation showed no significant influence on closeness. These findings suggest that black identity may be independent of feelings about whites and strongly related to one's immersion in black culture. The question of how childhood interracial contact and parental socialization affect identity remains unanswered.

With an understanding of the effect of childhood factors on adult black identity and assessment of the hypothesized processes by which they operate, one final question remains — what are the consequences for the individual and the black community of an increase or decrease in feelings of closeness to other blacks? This is a very controversial issue. There is a sense in which a strong racial identity can be seen as an advantage for blacks. Black communities are plagued by a disproportionate share of this country's poverty, violent crime, and unemployment. In addition to the problems of black communities, individual blacks of all socioeconomic statuses continue to face discrimination. Wealth may be able to reduce the frequency of discrimination in personal interactions, but in public places all blacks are susceptible to random acts of racial prejudice (Feagin 1991). Those who advocate the strengthening of black identity hope that it will be accompanied by efforts to eliminate racial prejudice and other structural barriers that stand in the way of all black Americans. A common example used by advocates of black identity is the modern civil-rights movement. It saw teachers, ministers, and entrepreneurs join forces with domestic workers, students, and sharecroppers to overthrow the forces of legalized discrimination (Morris 1984).

Despite this example of how identity can inspire positive changes, some prefer to point to the dysfunctional nature of black identity (Steele 1990). From their perspective, black identity is the barrier that has hurt blacks' ability to take

advantage of the new opportunities created by affirmative action, racial quotas, and other attempts at righting past wrongs. A frequent example used by this side to illustrate their beliefs about identity involves school performance. They cite studies showing a trend among blacks to see academic excellence as a betrayal of the race (Fordham & Ogbu 1986). As a result, fewer blacks are in a position to capitalize on the efforts of universities to recruit a diverse student population, and those who are accepted often do not obtain a college degree. The net effect of a strong black identity is inadequate job skills and a discouraging future. If blacks ignored the pressures of the group and concentrated on achieving personal success, the result would be a slow, steady climb of the race from its current position near the bottom of the stratification system.

In order to resolve the debate between these two sides, much work needs to be done on the *consequences* of black identity. Some behaviors that might be examined are interracial dating and marriage, involvement in social protests, participation in programs aimed at helping blacks, residence in black neighborhoods, academic achievement, and financial success. By assessing the impact of black identity on these and other outcomes a better understanding of its desirability for the individual and society might be achieved. However, this new understanding will be misleading unless other types of identity are simultaneously considered. Blacks feel a sense of allegiance not only to their race, but also to their genders, classes, regions, generations, religions, and other groups with whom they share traits. Attention to the determinants of these identities and their associated behaviors is crucial to an understanding of how allegiances affect behavior. Only with this knowledge can social scientists provide a complete assessment of the costs and benefits of integration for black children and the black community at large.

Notes

1. In discussing the concept of identity it is necessary to draw a distinction between personal and group identity. Personal identity refers to the feelings one attaches to a component of the self. Personal efficacy and self-esteem are two dimensions of this type of identity. Where personal identity is expressed as an "I," the group identity is a "we." According to Cooley (1902), "One identifies himself with a group and speaks of the common will, opinion, service, or the like in terms of 'we' and 'us'" (209). Gordon (1964) employs a similar definition of group identity in referring to ethnics as possessing "a shared feeling of peoplehood" (24). While it is true that people often possess comparable group and personal identities based on a single characteristic, this correlation should not be misinterpreted as implying the strengths of the two identities need be equal (Cross 1985). In other words, an individual's assertion that he or she is proud to be black in no way assures that the same individual will feel a bond with other African Americans.

2. This variable is identical to the construction of parental socialization developed by Demo & Hughes (1990). Readers interested in a more complete discussion of its construction should refer to this previous article.

3. The correlation between the two dimensions of outgroup orientation is .24.

4. Participation in a black church is employed as a measure of black culture. While this is admittedly only a part of the activities and belief structure traditionally associated with blacks in the U.S., research has shown it to be a central element of the community (DuBois [1899] 1973; Morris 1984).

5. In the SES coding scheme, managers and officials are categorized as professionals. Proprietors of unincorporated businesses are excluded from the analysis. They represent less than 2% of the sample. Proprietors of incorporated businesses are categorized according to their business.

6. All analyses were conducted with the number of cases fixed at 1,452. This represents the number of cases valid across all five equations and thus provides for a consistent case base. Separate analyses with the number of cases unrestricted did not significantly alter any of the substantive results reported here.

7. The absence of evidence for strong effects of interracial contact and parental socialization on the likelihood of regularly attending a traditional black church are attributable to the inclusion of a regional control variable in the equation. Unreported analyses excluding region produced strong and reliable estimates for all but the cautious/defensive parental messages. With region added to the model, the coefficients on other variables decline dramatically. An example is childhood interracial contact. Before including regional controls, the coefficient on contact was -.3280. In Table 2 it is only -.0935. That translates to a difference of more than 20 percentage points in the calculated effect of contact on church participation. Further testament to the power of region in this equation is its observed coefficient. Table 2 indicates that childhood residence in the South increases the likelihood of adult involvement with a traditionally black church by 141%. In this and other equations throughout the article, region shows itself to be a tremendously important factor in many areas of African Americans' lives.

References

Abrams, Dominic, and Michael A. Hogg. 1990. *Social Identity Theory*. Springer-Verlag.

Allen, Richard L., Michael C. Dawson, and Ronald E. Brown. 1989. "A Schema-Based Approach to Modeling an African-American Racial Belief System." *American Political Science Review* 83:421-41.

Blauner, Robert. 1972. *Racial Oppression in America*. Harper & Row.

Broman, Clifford L., James S. Jackson, and Harold W. Neighbors. 1989. "Sociocultural Context and Racial Group Identification among Black Adults." *Revue Internationale de Psychologie Sociale* 2:369-78.

Broman, Clifford L., Neighbors, Harold W., and Jackson, James S. 1988. "Racial Group Identification among Black Adults." *Social Forces* 67:146-58.

Cooley, Charles Horton. 1902. *Human Nature and the Social Order*. Scribner's.

Cross, William E., Jr. 1985. "Black Identity: Rediscovering the Distinction between Personal Identity and Reference Group Orientation." Pp. 155-71 in *Beginnings: The Social and Affective Development of Black Children*, edited by Margaret B. Spencer, Geraldine K. Brookins, and Walter R. Allen. Erlbaum.

Demo, David H., and Michael Hughes. 1990. "Socialization and Racial Identity among Black Adults." *Social Psychological Quarterly* 53:364-74.

DuBois, W. E. B. [1899] 1973. *The Philadelphia Negro*. Kraus Thomson Organization.

Ellison, Christopher G. 1991. "Identification and Separatism: Religious Involvement and Racial Orientations among Black Americans." *The Sociological Quarterly* 32:477-94.

Feagin, Joe R. 1991. "The Continuing Significance of Race: Antiblack Discrimination in Public Places." *American Sociological Review* 56:101-16.

Fordham, Signithia, and John U. Ogbu. 1986. "Black Students' School Success: Coping with the 'Burden of *Acting White*'." *The Urban Review* 18:176-206.

Gordon, Milton M. 1964. *Assimilation in American Life*. Oxford University Press.

Hughes, Michael, and David H. Demo. 1989. "Self-Perceptions of Black Americans: Self Esteem and Personal Efficacy." *American Journal of Sociology* 95:132-59.

Hughes, Michael, and Bradley R. Hertel. 1990. "The Significance of Color Remains: A Study of Life Chances, Mate Selection, and Ethnic Consciousness among Black Americans." *Social Forces* 68:1105-20.

Jackson, James S., and Gerald Gurin. 1987. *National Survey of Black Americans, 1979-80* (machine readable codebook). Inter-University Consortium for Political and Social Research, University of Michigan.

Jackson, James S., Walter R. McCullough, and Gerald Gurin. 1988. "Family, Socialization Environment, and Identity Development in Black Americans." Pp. 242-56 in *Black Families*. 2d ed., edited by Harriette Pipes McAdoo. Sage.

Jackson, James S., Walter R. McCullough, Gerald Gurin, and Clifford L. Broman. 1991. "Race Identity." Pp. 238-53 in *Life in Black America*, edited by James S. Jackson. Sage.

Morris, Aldon D. 1984. *The Origins of the Civil Rights Movement*. Free Press.

Park, Robert Ezra. 1950. *Race and Culture*. Free Press.

Phinney, Jean S. 1990. "Ethnic Identity in Adolescents and Adults: Review of Research." *Psychological Bulletin* 108:499-514.

Steele, Shelby. 1990. *The Content of Our Character: A New Vision of Race in America*. St. Martin's Press.

Stinchcombe, Arthur L. 1975. "Social Structure and Politics." Pp. 557-622 in *Handbook of Political Science*, edited by Nelson Polsby and Fred Greenstein. Addison-Wesley.

The Impact of Drug Use on Earnings: A Life-Span Perspective*

DENISE KANDEL, *Columbia University and New York State Psychiatric Institute*
KEVIN CHEN, *Columbia University*
ANDREW GILL, *California State University, Fullerton*

Abstract

Cross-sectional and longitudinal causal analyses were carried out on a longitudinal cohort of employed adult males (N~ 400) to investigate the relationship between various forms of drug use and earnings. By age 35, there are no cumulative effects of the use of illicit drugs on wages. Specification of the relationships at various stages of the life cycle reveals that the effects of illicit drug use are positive in the early stages of labor force participation (by the late twenties) but negative in later stages (by the mid-thirties). The negative impact of drug use on earnings appears to take more than a decade to manifest itself and is reflected in lower rates of earnings growth. We propose that a labor market theory of life-cycle compensation contracts and investments in training and the differential impact that different types of job changes have in the short and the long run among users and nonusers provide a partial accounting of the processes underlying the observed opposite relationships between drug use and earnings at different age periods. The importance of adopting a life-span perspective is highlighted by our findings.

Drug Use and Labor Force Achievement

Until recently, lifestyle factors, especially those indexing nonconforming values and behaviors, were rarely included in models of socioeconomic achievement. However, involvement in drugs, especially illicit drugs, may be an important factor to consider among the determinants of achievement.

** Work on this research has been partially supported by research grants DA03196, DA04866 and DA02867 (Denise Kandel, Principal Investigator) and a research scientist award DA00081 to Denise Kandel from the National Institute on Drug Abuse. Partial support for computer costs was provided by Mental Health Clinical Research Center Grant #MH30906 from NIMH to the New York State Psychiatric Institute. We are indebted to Mark Davies and to the anonymous reviewers of this journal for comments on an earlier version of this article. Direct correspondence to Denise Kandel, 722 West 168th Street, Unit 20, New York, NY 10032.*

Relatively little is known about the possible consequences of drug use on occupational careers and earnings in adulthood. Moreover, the effects of drug use on work roles and earnings documented to date are conflicting and somewhat puzzling. Illicit drug use has been shown to have negative effects on labor market activity, as reflected in increased job mobility (Kandel & Yamaguchi 1987; Newcomb 1988), number of missed work days (Robbins, Martin & Kaplan 1985; Zarkin, French & Rachal 1992), rate of unemployment, length of time unemployed (Kandel 1984; Kandel & Davies 1990; Miller, Cisin & Harrell 1978; O'Donnell et al. 1976; White, Aidala & Zablocki 1988), and reduced rates of promotion (Johnson & Herring 1989). Paradoxically, drug use has been found either to be unrelated to earnings (Jessor, Donovan & Costa 1991; Kaestner 1994; White, Aidala & Zablocki 1988) or to be related to higher wages. This positive relationship has been observed in several samples, including a cohort of New York State adolescents that we have followed into adulthood (Kandel, Mossel & Kaestner 1987), a cohort of former high school students in California (Newcomb 1988), and a large national cohort of young Americans, the National Longitudinal Survey of the Labor Force Experience of Youth, the National Longitudinal Survey of Youth (NLSY) (Gill & Michaels 1992; Kandel & Davies 1990; Kaestner 1991 1994; Mijares 1992; Register & Williams 1992). Some of the effects are modest in size.

Drug use could be expected to be associated with lower earnings through lower levels of educational attainment, less total work experience, poorer performance on the job, or lower commitment to traditional criteria of success. We noted above the consistent relationships observed between illicit drug use and job turnover. Illicit drug use is also a determinant of dropping out of high school (Mensch & Kandel 1988). And, despite the high prevalence of the use of illicit drugs in our society, young people who use drugs are usually less conforming and less attached to social institutions than nonusers (Kandel 1984; Johnston, Bachman & O'Malley 1984; Newcomb 1988). Illicit drug use has been shown to have negative consequences on the timing and quality of participation in adult social roles. Thus, drug use is related to a delay in the entry into marriage and in parenting, and to an increase in marital separation if drug use continues after marriage (Bachman, O'Malley & Johnston 1984; Newcomb & Bentler 1988; Yamaguchi & Kandel 1985a, 1985b). Drug users may have reduced aspirations and lower commitment to values underlying the striving for success. Drug use may also impair functioning on the job. The users' values and behaviors may send out unfavorable signals to employers, thereby hindering advancement on the job. Since drug use has been found to have negative consequences on dimensions of labor force participation that are positively related to wages, the observed null or positive relationships with wages are somewhat surprising and have puzzled researchers.

We here propose and document that a life-span perspective may help resolve the puzzle. The impact of drug use on earnings may vary depending upon whether the workers are in an early or a later stage of their careers. Indeed, Mullahy and Sindelar (1991, 1993), relying on diagnostic criteria specified by the American Psychiatric Association, found with respect to alcoholism no effect on income among 22 to 29 year olds but a negative effect among those 30 to 44 and 45 to 59. Kaestner (1994), using the 1984 and 1988

waves of the NLSY, estimated a fixed-effects model to control for unobserved heterogeneity in the sample, and found that the positive effect of illicit drug use on wages among NLSY men found in previous studies became negative, but statistically insignificant.

Analyses completed to date by ourselves and by others on the role of drug use on economic outcomes have been carried out only on young workers. Our own earlier analyses on a New York State cohort were carried out when respondents were 28 to 29 years old. Analyses on the NLSY are based on respondents at ages 18 to 31. We speculated at the time we reported our results that a potential negative impact of drug use on earnings may be lagged and may not manifest itself until later in a drug user's work career (Kandel, Mossel & Kaestner 1987). Similarly, Gill and Michaels (1992) proposed that drug users would eventually experience lower rates of wage growth than nonusers because rates of returns to working experience and education were lower for drug users in their mid-twenties.

This study delineates the effects of drug use on wage levels and wage growth by age groups and tracks the relationship between drug use and earnings among men from early adulthood to their mid-thirties.

Theoretical Background

DETERMINANTS OF EARNINGS

Earnings are primary indicators of socioeconomic achievement associated with a job. The human capital model constitutes the most basic model of wage determination. Among individual determinants of earnings, educational attainment and labor force experience are the most important positive factors (Blau & Duncan 1967; Featherman & Hauser 1978; Griliches & Mason 1973; Jencks et al. 1979; Mincer 1974; Willis 1986), with health status somewhat less important (Grossman 1972; Mullahy & Sindelar 1993). By contrast, the consequences of job mobility for earnings and earnings growth are unclear. Topel and Ward (1992) document that job changes are a major determinant of earnings of young people, accounting for at least a third of early-career growth and ultimately leading to the stable employment associated with mature careers. But job mobility may also reduce earnings (Topel 1991). The job search paradigm stresses that certain job shifts early in a career represent attempts by young workers to improve the match between their own needs and aspirations and the characteristics and rewards of a job. However, not all young workers will follow this pattern. Individuals are likely to differ for unobservable reasons in the types of voluntary changes they make. The negative consequences of job mobility can also arise if the prospect of changing jobs creates a disincentive for workers to invest in job-specific training (Borjas 1981).

Several years of labor force experience are required to identify the effect of education on earnings because job earnings immediately after completion of schooling do not adequately reflect the full returns of educational investment, even after controlling for the level of experience of workers who have completed school earlier (Griliches 1976, 1980; Mincer 1974). The best time to measure the effect of schooling on earnings is at the point of overtaking, after about eight

years in the labor force, when the returns to additional schooling are about equal to the opportunity costs of further investments in schooling (Mincer 1974). Not only education, but ability and background characteristics also appear to have varying impacts on earnings at different points in a career (Blau & Duncan 1967; Hauser & Daymont 1977). Wages in the first two years bear almost no relation to wages three to four years later (Meyer & Wise 1982).

A perspective that considers that the determinants of earnings may vary over the life of an individual would seem important.

VOLUNTARY AND INVOLUNTARY JOB MOBILITY

Of the relationships that have been observed between illicit drug use and characteristics of a work career, the relationships with job turnover and unemployment are the strongest and the most consistent (e.g., Kandel & Yamaguchi 1987). However, to the best of our knowledge, no investigation has examined such turnover as a function of whether it is voluntary, such as leaving for a better job, or involuntary, such as being fired. Although this distinction is somewhat difficult to sustain, for being fired may result from a worker's willful inappropriate behavior at work, a consideration of the nature of job turnover may be helpful in understanding the relationship between illicit drug use and earnings.

Voluntary and involuntary job changes may have different consequences for a person's earnings, whether short-term or long-term. Involuntary leaves have generally been found to be associated with lower earnings (Blank 1989). Voluntary job leaves can lead to unemployment. Some of these leaves, such as those due to a demotion or wage cuts, may have the same negative psychological impact as the experience of an involuntary job leave. Nevertheless, because the decision as to whether or not to leave lies with the individual, voluntary job leaves are distinct from involuntary leaves. The rates of involuntary job turnover will depend on characteristics of jobs and firms as well as of persons. We are particularly interested in individual characteristics that may increase the risk of such job turnover, and especially those related to the use of illicit drugs. Indeed, personal nonconformity that is reflected in the use of drugs may be an important determinant of involuntary job leaves, and as such may be a potential explanatory factor in the relationship between drug use and earnings.

LIFE-CYCLE PERSPECTIVE ON WAGE GROWTH

The controversial positive effect of drug use on earnings has puzzled researchers. We suggest that labor market theories of life-cycle compensation contracts and investments in training may help interpret the findings.

As noted by Lazear (1979), workers care about the present value of their own lifetime earnings, while firms are concerned about the present value of their workers' productivity. However, close supervision of workers is costly to firms, making it difficult to monitor and verify productivity (Carmichael 1989). One option for the firm is to sequence pay, whereby workers are paid a relatively low wage early in their tenure and a wage higher than they could get elsewhere based on their marginal productivity later in their careers to

compensate (Lazear 1979). By deferring payment a firm may induce a worker to perform at a higher level of effort. Workers are less likely to shirk because dismissal means surrendering the deferred compensation. This type of compensation scheme will appeal most to individuals with longterm goals who plan to work hard and long enough to get the deferred payment. Jobs that are long-term in nature are also more likely to be those that provide large firm-specific investments in training. The consequences of this process for wage growth over the life cycle have been emphasized by Mincer (1974) and by Borjas (1981). Borjas, for example, shows that job mobility leads to wage gains in the short run, but that nonmobile workers tend to achieve significantly higher wages over the long run (Borjas 1981). Job mobility is thought to affect the slope of the experience-earnings profile, with earnings growing at a slower rate on jobs that are not expected to be permanent. As Borjas notes, the prospect of labor turnover will create disincentives for investing in on-the-job training.

Two outcomes can be predicted based on this life-cycle perspective. First, young workers who opt for jobs that sequence pay or provide training will earn less early in their careers than comparable workers who do not make such investments. Second, earnings growth and, at some point, earnings will be higher for those who take these career paths than for others. These workers capture the deferred payment from wage sequencing and they share in the productivity gains resulting from the firm-specific investments.

Drug use may be differentially associated with the less favorable career path. As mentioned, drug users may have reduced aspirations and lower commitment to values underlying the striving for success. They have also been found to change jobs more frequently than nonusers (Kandel & Yamaguchi 1987). Thus, users in their twenties, at the beginning of their work career, may take jobs that are less likely to require a long-term commitment and provide much training. Nonusers, on the other hand, may accept a relatively low wage so as to qualify for jobs that offer long-term employment and in the future will pay higher wages than the market wage being earned by others. While the wage-sequencing and training hypotheses do not predict when the earnings of nonusers will overtake the earnings of users, they do suggest that nonusers will have higher rates of earnings growth.

We here report the results of analyses based on data obtained from a longitudinal cohort, when the members of the cohort were in their mid-thirties. We investigate the effects of different classes of drugs, cigarettes, alcohol, but especially the use of marijuana and other illicit drugs, on earnings. Our goals are to identify the potential impact of drug use on wages in different phases of workers' careers, as workers age from their late twenties to their mid-thirties, and to gain some understanding of the underlying mechanisms, in particular the potential role of labor market theory of wage compensation and of job turnover in accounting for the relationship between illicit drug use and wages. To parallel previous work in the field, we examine the impact of cumulative drug use at two specific ages. We then examine the impact of drug use on rates of earning growth at two different periods of the life span.

Data and Methods

We carried out cross-sectional and longitudinal analyses on members of a longitudinal cohort that we have followed for 19 years, from age 15-16 to age 34-35.

SAMPLE

The analyses are based on data for adult men derived from a follow-up cohort of men and women ($N = 1{,}160$), last interviewed at age 34-35. The cohort is representative of adolescents formerly enrolled in grades 10 and 11 in public secondary schools in New York State in 1971-72. The original two-stage sampling procedure involved the selection of a stratified sample of 18 high schools and a sample of students clustered by homerooms and stratified to represent the different grades within each school. Students who had not participated in the initial study and presumably were chronic absentees were also selected (and sampled at a lower rate) to permit unbiased estimates of the former student population at the time of the adult follow-up. Subsequent to the initial survey, three adult follow-ups were carried out in 1980, 1984, and 1990.[1] Overall, 71.7% of the initial adolescent target sample still alive was reinterviewed in 1990.

As we had found in prior surveys (Kandel, Raveis & Logan 1983), there were few statistically significant differences between participants and nonparticipants in the 1990 survey. However, compared with the initial high school sample, the reinterviewed sample reflected greater attrition of minorities than of whites: the sample includes 86.7% whites in 1990 compared with 79.6% initially.

Data were obtained through structured personal household interviews, which took on the average 90 minutes to administer. The schedule consisted almost exclusively of structured items with closed-end response alternatives and two charts designed to reconstruct the respondent's life and drug histories on a monthly basis. The drug histories were obtained from respondents who reported having used each drug ten times or more in their lives, so as to exclude experimental users. Information collected on the respondents' employment histories includes start and stop dates, part-time or full-time status, salary level, and reason for leaving the job for each job held.

Sampling weights took into account all relevant features of the sampling design. Adjustments for the differential participation of the target follow-up sample in 1980, 1984, and 1990 were also included in the weights.

ANALYTICAL STRATEGIES

Three major analyses and two variants were carried out on the basis of different statistical strategies, different subsamples, and slightly different definitions of the predictor variables. These strategies enabled us to move from considering a wide range of variables cross-sectionally to an increasingly smaller set of factors dynamically and to deal with three substantive issues.

(1) To predict levels of earnings, multiple regression analysis was conducted among all men working full time at two points in time: (a) in 1984, at age 28-29 ($N = 419$); and (b) in 1990, at age 34-35 ($N = 439$).

The specification of the single-equation model follows that of most human capital models (Mincer 1974; Mincer & Jovanavic 1981; Willis 1986). The dependent variable is the log of annual earnings; the explanatory variables include socioeconomic background variables, lifestyle and school performance measures ascertained in the initial high school survey (grades, degree of peer activity); total labor force experience, tenure on the current job, number of hours working, and individual characteristics at the time of the interview that could be related to both drug use and wages (e.g., education, marital status, health status and delinquency); and drug use. Tenure was entered together with total labor force experience in order to differentiate returns to employer-specific skills and training from those associated with general labor force experience (Altonji & Shakotko 1985; Mincer & Jovanovic 1981; Topel 1991). Labor force experience and tenure were also entered as quadratic terms reflecting the potentially diminishing returns to on-the-job training investments.[2]

The time-varying covariates were measured *cumulatively* up to each survey date, i.e., 1984 and 1990.

(2) To investigate the predictors of period-specific changes in earnings, two analyses were implemented: multiple regression and causal modeling.

(a) A variant of the multivariate regression model described above was estimated, where two classes of time-varying predictors, the drug-use variables and number of job changes, were measured *for the intervals* between the prior and the current interviews instead of *cumulatively*. The two intervals covered the four-year period from 1980 to 1984 and the six-year period between 1984 and 1990. Because earnings at the prior interview were also included as a predictor in the equations, this strategy enabled us to observe the specific effects of drug use and job turnover on earnings growth at different periods of a man's work career, from the mid- to the late twenties, and from the late twenties to the mid-thirties. This analysis allowed us to include a large set of covariates.

(b) A three-wave causal model was estimated on men who were working full time in 1980, at age 24-25, and who were still working (whether full time or part time) when reinterviewed in 1984 and 1990, at ages 28-29 and 34-35, respectively ($N = 401$). This multiwave longitudinal causal model permitted us to examine the potential reciprocal and dynamic relationships between drug use, job turnover, and earnings. The model, which includes a subset of the variables included in the multiple regressions, was estimated through structural equation modeling, LISREL 7 (Jöreskog & Sörbom 1989).

(3) To test a hypothesis suggested by the findings in step 2 above, variants of analysis 1b and analysis 2b were implemented. (a) An additional multiple regression was estimated on full-time workers at age 34-35, with all variables measured cumulatively up to age 34-35. The model was identical to that estimated in (1-b) above, except that the sample was restricted to those men who did not use any illicit drugs between ages 28-29 and 34-35 ($N = 347$) and earnings at age 28-29 were also included in the model. (b) As a final step, the three-wave causal model in analysis 2b was estimated separately for men who had not used any illicit drugs 10 times or more in their lives and those who

had. The multigroup comparison feature of LISREL (Jöreskog & Sörbom 1989) was used to estimate the equality of paths between the groups.

METHODOLOGICAL CONSIDERATIONS

Sample Selectivity

The different subsamples of men in the labor force vary in size in each of the three survey dates, 1980, 1984, or 1990. A number of respondents were not working at each survey. Sample selectivity may bias the estimates of wage levels if the likelihood of working is determined by unmeasured factors similar to those that determine earnings (Fligstein & Wolf 1978; Griliches, Hall & Hausman 1978; Heckman 1980; Mare, Winship & Kubitschek 1984). Inconsistent findings have been reported regarding the statistical significance of a sample selection correction term for employment in prior research linking drug use to wages. The term was found not to be significant by Kandel and Davies (1990) and Kaestner (1991), but to be significant by Gill and Michaels (1992) and Register and Williams (1992). In the linear models that we estimated, we included a correction term adjusting for the potential effects of differential propensity to work full-time, using the two-step procedure suggested by Heckman (1980). A probit model was run on the total sample of men to predict the likelihood of employment.[3] A correction term, the inverse Mill's ratio (λ) obtained from the probit model, was entered into the regression models. We calculated the correction terms with a set of parallel variables for both the 1984 and 1990 samples (see Appendix C).

In addition to its theoretical interest, the inclusion of job turnover in the models controls for other types of person-specific heterogeneity. The inclusion of labor force experience and tenure on the job raises additional methodological considerations related to unobserved heterogeneity. If there are unmeasured person and/or job-specific components of wages, tenure and/or experience, the latter variables and the error term in the wage equation would be correlated (Altonji & Shakotko 1985; Mincer & Jovanovic 1981). These correlations could bias estimates of all the coefficients, including the drug-use variables. The inclusion of measures of prior job mobility in the model controls for such person-specific heterogeneity (Mincer & Jovanovic 1981).

Specification of Drug Effects

A basic issue underlying any analysis of the impact of drug use on individuals is the extent to which the effects observed are due to the characteristics of individuals who self-select themselves into drug use or to the effects of using the drugs. The drug effects can be physiological as well as social and psychological. Being a drug user may lead one to remain embedded in a particular social context, which would sustain nonconforming lifestyles, including nonparticipation in expected adult roles. In the absence of experimental laboratory and field data, it is virtually impossible to separate out these two processes empirically. In the models we have estimated, in order to control for a possible selection effect in assessing the effects of drug use on the outcome being predicted, we controlled for factors that may be related to the onset of drug use, such as sociodemographic characteristics, behaviors measured in adolescence, education-

al attainment, other forms of deviance reflected in delinquent participation, and levels of adult social role participation, including current marital status and accumulated work experience both in the sample selection and the wage models.[4]

Other Statistical Considerations

In order to maximize the sample sizes, those who were absent from school at the time of the 1971 survey and for whom adolescent measures were not available were included in the analysis. A dummy missing variable was included in the models to account for missing data on the covariates measured in the 1971 survey.

A larger number of factors were entered in the multiple regressions than in the causal model. In order to reduce the complexity of the three-wave causal model, a subset of variables were selected, either because they were statistically significant in the regressions or because of their theoretical interest. As noted above, the time interval covered by the variables, whose value could change over time, differed in different models. In the first set of regressions, these variables, such as number of prior job changes or number of years a particular drug class had been used, were counted cumulatively from 1971 up to each interview. In the second set of regressions and in the causal model, these variables were counted cumulatively from the beginning to the end of each interval, i.e., between the 1980 and 1984 waves, and the 1984 and 1990 waves; in the causal model, a count was also made from 1971 to the first follow-up interview in 1980. The intervals covered 13 or 19 years in the first set of regressions, four and six years in the second set, and nine, four, and six years in the causal model.

DEFINITION OF VARIABLES

To be included in the analyses, a respondent had to be employed full-time at the time of the interviews included in each respective analysis, except for the causal model, where full-time work status was required only as of 1980, and part-time employment was acceptable for subsequent waves in 1984 and/or 1990. Of those employed full-time in 1980, 3.3% were employed part-time by 1984, and 2.6% by 1990. Employment at interview was defined as holding a civilian nonfarm job during the interview month, or one month before or after the interview if temporarily out of work at the time of the interview.

The dependent variable in all the regression analyses was a log measure of earnings, measured at the time of each survey. Current wages could be reported in the form of weekly, monthly, or annual earnings, with 14 equivalent precoded categories listed under each time period. The earnings were inputted at the midpoint of each category, and the annual equivalents ranging from $2,000 to $85,000 used in the regressions. The logarithm of earnings is the preferred functional form on theoretical and empirical grounds (Heckman & Polachek 1974; Mincer 1974), with the added advantage that the regression coefficients can be interpreted approximately as the percentage change in earnings due to a unit change in the covariate.

The following covariates were entered in the multiple regressions:

Background and Individual Characteristics (at the Time of Each Interview)

Race (black = 1, nonblack = 0)

Highest year of schooling completed by parents

Highest year of schooling completed

Current marital status (1 = married, 0 = others)

Self-assessed health status (1 = poor to 4 = excellent)

Delinquency index within the last 12 months, which counts the number of acts committed out of a list of 15 (scores range from 0 to 11) (α = .65)

Adolescent Characteristics (Measured in the 1971 High School Survey)

Peer activity index based on the extent of participation in any of five activities involving peers, i.e., getting together or hanging out with friends after school, dating, attending parties, or driving around [scores range from 10 (low) to 40 (high)] (α = .69).

Dummy missing on peer activity to account for those absent from school in the 1971 survey (27%) who are missing information on the adolescent variables

Self-reported grade point average (GPA). For school absentees, the value was imputed by regression from retrospective reports provided in the 1980 survey.

Characteristics of Labor Force Participation

For characteristics that can change over time, such as number of job changes, the time interval over which the variable was measured varied in different models.

Labor force experience (in years) was measured as of the first full-time nonfarm civilian job, up to 1984 or 1990 (depending on the model).

Labor force experience squared (years)

Job tenure: number of years in current job

Job tenure squared (years)

Average hours worked per week

Number and type of job changes. Each job termination was differentiated according to whether it was voluntary (e.g, getting a better job), involuntary (fired, layoff, temporary job), or other (family responsibility, illness). As we noted earlier, the conceptual distinctions among these categories are somewhat arbitrary. In the causal model, the last two classes were combined into a nonvoluntary category. The number of job changes consisted of the cumulative number of all full-time and part-time employment spells with different employers as of 1971 up to each interview date in the first set of regressions; the number during each of two different time intervals, between 1980 and 1984 and between 1984 and 1990, for the second set of regressions; the number during these two intervals and the additional period between 1971 and 1980 for the causal model.

Drug Use Variables

The drug use variables were defined so as to capture fairly regular and sustained levels of drug use. Because of differences in the use patterns of different classes of drugs and levels of involvement that these patterns represent, different criteria were used for the four classes of drugs under study. The cumulative number of years that drugs were used was obtained by dividing by 12 all months a particular class of drugs was used (at the specified intensity) since the 1971 survey. For the two specific periods in adulthood, the counts were restricted to the four-year interval between 1980 and 1984 and the six-year interval between 1984 and 1990.

Years of cigarettes smoking half pack a day or more

Years of using alcohol 4 or more times a week (near daily)

Years of using marijuana at least once a month

Years of using illicit drugs other than marijuana (cocaine, psychedelics, heroin and use without a prescription of minor tranquilizers, stimulants and sedatives) at least once a month. For the causal model, a latent factor for illicit drug use was estimated, with the manifest indicators being the two illicit drug use variables (number of years of use of marijuana and number of years of use of other illicit drugs). These cumulative drug use variables are highly correlated with an individual's severity of drug involvement. The longer the period of use, the greater the involvement. The correlations between total number of months a particular drug class has been used by each survey date and the frequency of use of that drug within the last year preceding the survey range from .66 to .86.

The detailed data on drug use and employment histories made it possible to define drug use patterns and number of job changes at different life stages: cumulatively by the late twenties and by the mid-thirties; and for specific periods of the life span. All time-varying covariates were evaluated as of the time of the current interview. As indicated below, different start dates were used in different models.

Sample Selection Correction Term

A term to correct for a potential bias in the earning equations due to selection into employment was calculated as per the procedure outlined above and entered into the regression models.

Means and standard deviations for the variables measured among full-time workers as of 1984 and 1990 are presented in Appendix A. On average, the men have completed two years of college and have worked eight years by 1984 and 13 years by 1990. They have held three jobs by their late twenties and five by their mid-thirties. Job leaves are evenly divided between voluntary and nonvoluntary types.

Results

As background to the analysis, it is useful to consider the extent of drug involvement by the men in the cohort. (See Table 1.) Most have experimented with alcohol, cigarettes and marijuana in their lives, and more than half have also tried illicit drugs other than marijuana. However, the proportions who are currently using illicit drugs decreases dramatically from the mid-twenties to the mid-thirties. While close to 80% of the men had used marijuana, by age 34-35, less than 20% reported using it within the last year. At that age, very few men were still using illicit drugs other than marijuana.

PREDICTORS OF WAGES AT TWO DIFFERENT TIME PERIODS IN THE LIFE-SPAN

As a first step, to replicate the analytical strategy used in prior research, we estimated multivariate regression models separately for full time workers who were 28-29 years old (in 1984) and 34-35 years old (in 1990). The results are presented in Appendix B. (The correlation matrixes are available from the authors on request.)

The cumulative impact of drug use on earnings depends upon the stage in a worker's life. By the late twenties, the cumulative impact of marijuana use over a 13-year period from adolescence to young adulthood is positive, while the effect of cigarette smoking is negative, controlling for significant human capital and background variables and correcting for sample selection bias. However, by the middle thirties, there is no apparent effect of cumulative illicit drug use over the 19-year interval since adolescence. The regression coefficients for use of marijuana and of other illicit drugs are zero. Illicit drug use from the late twenties to the mid-thirties may have a different relationship with earnings than use at a younger age.

Job mobility has a different impact on earnings, depending upon whether mobility is voluntary or nonvoluntary, and depending upon the man's age. Cumulatively, nonvoluntary changes lead to lower earnings by the late twenties, and voluntary changes lead to higher earnings by the mid-thirties. Voluntary changes have no impact on earnings at the younger ages and involuntary and other changes retain a slight negative impact at the later ages. The role of variables other than drug use on earnings is consonant with the human capital model. Controlling for all other covariates, educational attainment is consistently the most important predictor of earnings. The other predictors are model-specific. Total labor force experience, extent of peer activity as an adolescent, and good health have a positive impact up to the late twenties only. Longer working hours are associated with higher wages by the mid-thirties only. Delinquent participation has no effect at either time period.

The negative coefficient for the sample selection correction terms is statistically significant and implies that whether or not a man was employed full time is not a random process. If employed, individuals not included in the model might have lower earnings than those included (Dolton & Makepeace 1987).

TABLE 1: Drug Use at Different Periods among Men Who Worked Full Time in 1980 and Were Still Working in 1984 and 1990

	Ever Used by 1990 Age 34-35 (Percent)	Used Past Year in 1980 Age 24-25 (Percent)	Used Past Year in 1984 Age 28-29 (Percent)	Used Past Year in 1990 Age 34-35 (Percent)
Alcohol	99.2	94.4	92.9	91.2
Cigarettes	79.9	51.2	49.2	43.7
Marijuana	79.0	53.0	38.8	17.9
Any other illicit drug[a]	55.5	20.0	18.8	8.3

(N = 429)

[a] Any illicit drugs other than marijuana, including cocaine, psychedelics, heroin, nonmedical use of minor tranquilizers, sedatives, and stimulants.

PERIOD-SPECIFIC PREDICTORS OF CHANGES IN EARNINGS

The absence of cumulative effects of drug use by the mid-thirties compared with the positive effect observed by the late twenties suggests that the eventual rate of earnings growth might become lower for drug users than for nonusers.

Two-Wave Multiple Regressions

As a first step, in order to identify the effects of drug use within a specific period of the life span, we estimated multivariate regression models at two different age periods: (1) from the mid- to the late twenties, and (2) from the late twenties to the mid-thirties. The number of job changes and the years of drug use were measured for each specific time interval. The other covariates were identical to those entered in the first regressions described above. These models estimate the impact of drug use on *changes* in earnings, i.e., wage growth, from 1980 to 1984 and from 1984 to 1990, since wage levels at the beginning of each interval were controlled. (See Table 2.)

Differences in the impact of drug use on earnings are clarified when the analyses are restricted to specific age periods and time intervals (Table 2). From the mid- to the late twenties, controlling for earlier wage levels, the use of illicit drugs other than marijuana has a slight *positive* effect on wage growth. However, between the late twenties to the mid-thirties the use of illicit drugs other than marijuana has a statistically significant *negative* effect on wage growth. This negative effect rate in the six years preceding the interview at age 34-35 is stronger than the positive effect in the four year period preceding the interview at age 28-29.

Not surprisingly, prior earnings is the strongest predictor of subsequent wage growth. The impact of involuntary changes is negative but only during

TABLE 2: Parameter Estimates from Multiple Regressions Predicting Log Annual Salary at Ages 28-29 (1984) and 34-35 (1990) among Male Workers with a Full-Time Civilian Nonfarm Job[a]

	From Age 24-25 (1980) to Age 28-29 (1984)			From Age 28-29 (1984) to Age 34-35 (1990)		
Covariates	r	Model 1[d]	Model 2[d]	r	Model 1[d]	Model 2[d]
Background						
Parents' highest schooling	.123**	.002	.004	.102*	.000	.000
Race (1 = black, 0 = nonblack)	–.080+	–.031	–.043	–.155**	–.119	–.018
1971 High school variables						
Peer activity index (10 to 40)[b]	.131**	.015***	.014**	.008	-.002	-.001
Grade average (1 = F . . . 5 = A)	.115*	–.008	–.015	.170***	–.022	–.030
Individual characteristics (at time of interview)						
Married (1 = Yes)	.159**	.131**	.128*	.130**	.034	-.008
Total years of schooling	.249***	.067***	.091***	.416***	.058***	.061***
Self-rating of health status (1-4)	.242***	.098**	.076+	.107*	-.022	-.035
Delinquency within last year	.055	.025	.035	-.042	.003	.003
Work related (up to interview)						
Years of work experience	–.094*	.145+	.139+	–.197***	–.027	–.030
Squared years work experience	–.070+	–.008	–.008+	–.196***	.001	.001
Years in current job	.025	.029	.032	.027	-.004	-.002
Squared years in job	-.030	–.003	–.003	.006	–.000	–.000
Avg. hours worked per week	.011	-.000	-.000	.221***	.005**	.005**
No. of voluntary job changes[c]	-.006	.033	.035	.005	.030	.034
No. of involuntary job changes[c]	–.222***	–.128***	–.126***	–.075+	-.032	-.030
No. of other job changes[c]	–.155***	–.071+	–.071+	–.107*	–.144***	–.151***
Log job income at previous interview	.467***	.461***	.459***	.576***	.432***	.425***
Drug use (between prior and current interview in years)						
Cigarette smoking (1/2 pack a day)	–.142**	–.029+	–.027+	–.065+	.019*	.016+
Daily alcohol use	–.041	–.015	–.022	–.044	–.010	–.009
Monthly marijuana use	.014	.006	.006	–.099*	-.004	-.002
Monthly other illicit drugs use	.049	.033	.041+	–.123**	–.071*	–.063*
Sample selection correction (λ)	—	—	–.561	—	—	–.455*
Constant		–.366	-.355		1.562**	1.802***
Adjusted R^2		.364	.368		.439	.450
Total N		369			409	

[a] Life-stage specific work and drug use included in models.
[b] Missing dummy for peer activity included in the model.
[c] Between prior and current interview: from 1980 to 1984, from 1984 to 1990.
[d] Model 1, uncorrected estimates; model 2, estimates with correction term for sample selection.

+ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

the younger age period. In contrast to the cumulative models, voluntary changes have no significant impact on wage growth within each period and other types of job changes have a negative impact. Being married becomes significant (positive) from the mid- to the late twenties.

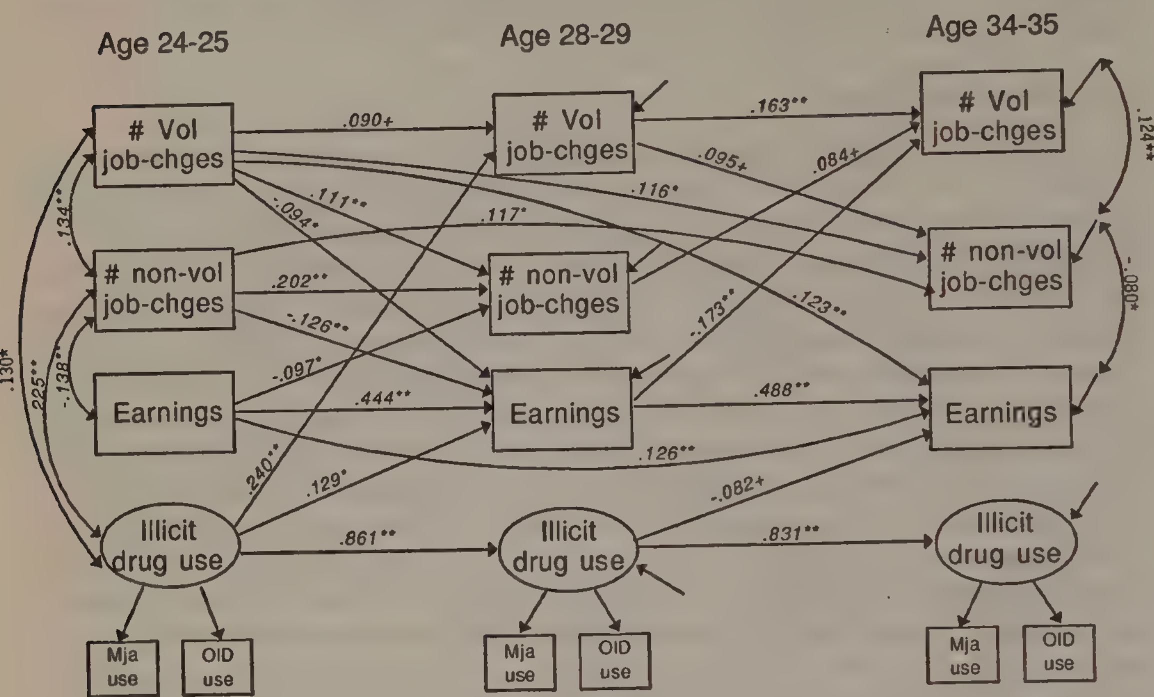
Three-Wave Causal Model

The three-wave structural equation model estimated among men, who had a full-time job at age 24-25 and were still working at ages 28-29 and 34-35 (Figure 1), provides a more powerful test of the relative impact of drug use on labor force performance over time than the multiple regressions. It also makes it possible to analyze the reciprocal and dynamic effects of drug use, earnings and job turnover on each other. Because of estimation constraints, a smaller number of variables were included in the causal model than in the regressions. The same variables were included in all three waves, except for race. Involuntary and other job changes were combined into a single variable, nonvoluntary job changes. To control for measurement error, we introduced a latent construct for the use of illicit drugs and allowed the error terms to correlate over time. Coefficients for job mobility, earnings, and illicit drug use are displayed in Figure 1. Race, years of schooling, self-reported health status and marital status were also controlled, but their coefficients are not shown. Since delinquency had no significant effect in the multiple regressions and its exclusion would not lead to a misspecification of the drug effects, it was not included in the causal model. Standardized coefficients are displayed in the figure.

The positive impact of illicit drug use on subsequent wage growth in the twenties and its negative impact in the thirties is confirmed. A reestimation of the model constraining the paths from drug use to income during each interval to be equal to each other indicates that the paths are significantly different from each other ($p<.01$). (The chi-square is 145.46 for the unconstrained model, 154.17 for the constrained model). By contrast, there are no significant direct or indirect effects from earnings to drug use. Controlling for initial levels of job turnover, illicit drug use predicts increased voluntary job mobility in the early stage of the work career, but does not predict either voluntary or nonvoluntary mobility in the later stage. The three-wave model clarifies the results regarding the impact of voluntary job changes by age 34-35 obtained in the cumulative and period-specific multiple regressions. Job mobility, but especially nonvoluntary job changes, early in a career (by age 24-25) have negative effects on subsequent wages earned in the late twenties; job mobility in the late twenties, whether voluntary or nonvoluntary, has no direct effect on wages six years later. However, voluntary job changes have a time-lagged positive effect on wages over the ten year period from the mid-twenties to the mid-thirties. Level of earnings in the mid-twenties (1980) reduces subsequent nonvoluntary job turnover while earnings in the late twenties reduces subsequent voluntary turnover.

Earnings and nonvoluntary job changes each have a time-lagged effect on themselves over the ten-year period, over and beyond the intermediate direct effects that they have from one period to another during that interval. We cannot, however, eliminate the possibility that other variables may have affected the same measure at different times. There is another time-lagged effect: earlier

FIGURE 1: Three-Wave Causal Model of Drug Use, Job Changes and Earnings[a]



$\chi^2 = 145.5$, df = 121, p = .064, AGFI = .93

[a] Male full-time workers at age 24-25, working at age 28-29 and 34-35. Standardized coefficients are shown. Race, year of schooling, self-reported health, and marital status are controlled for, but not shown. Nonsignificant paths, correlated measurement errors of drug use over the three surveys, and error correlations at age 28-29 are also not shown. Error correlation of earnings and nonvoluntary job changes is -.160**; correlation of nonvoluntary and voluntary job changes is .115*.
(N = 401).

+ $p < .10$ * $p < .05$ ** $p < .01$

voluntary job changes predict nonvoluntary changes. In fact, one type of job turnover also predicts subsequent participation in the other: voluntary job turnover in the mid-twenties is related to an increase in subsequent nonvoluntary turnover four years later, which in turn increases the risk of voluntary turnover in the following six years.

DIFFERENTIAL EFFECTS OF DRUG USE ON WAGE GROWTH: AN INTERPRETATION

The negative effect of illicit drug use on wage growth in the thirties is consistent with the notion that drug use lowers productivity; the positive effect in the twenties is controversial. As we noted in the introduction, labor market theories of life-cycle compensation contracts and investments in training may help

interpret and reconcile these period-specific findings. The hypothesis is that drug users early in their careers take jobs with less career potential and eventually their rate of earnings will decline compared to that of nondrug users. In a preliminary test of the first part of the hypothesis, we examined the types of jobs held by drug users and nonusers at age 24-25 in terms of their advancement potential. We attached to each job a code reflecting the training requirement of that job, as indexed by the specific vocational preparation defined by the Department of Labor (U.S. Dept. of Labor 1991:1009). This code reflects "specific vocational preparation" (SVP), the amount of time required to learn the techniques needed for average performance in a specific job-worker situation, and reflects vocational education, apprentice training, in-plant training, on-the-job training, and essential experience in other jobs (U.S. Dept. of Labor 1991:1009). There is a statistically significant difference between the SVP scores of drug users and nonusers, with the users scoring lower than the nonusers. This supports our hypothesis.

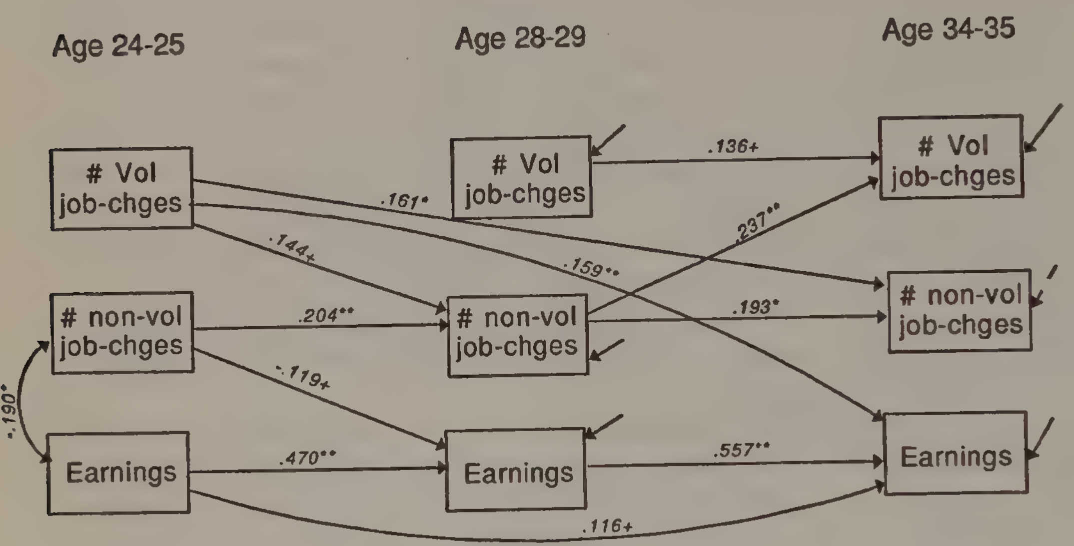
The second part of the hypothesis led to a reformulation of the multiple regressions and a specification of the three-wave model presented above in Figure 1. In a first step, we compared drug users who had stopped using drugs by their late twenties with persons who had never used any illicit drugs. We restricted the age 34-35 sample to those individuals who did not use drugs between ages 28-29 and 34-35 in order to isolate the time-lagged effects of earlier drug use on earnings growth from the negative period-specific effects of use documented above. The change in earnings over the six-year interval was regressed on the same predictors as those included in the multiple regression (see Appendix B, model 2), with the exception that cumulative drug use was measured up to age 28-29 (1984). The other individual and work related characteristics were measured up to age 34-35 (1990). Cumulative monthly use of illicit drugs other than marijuana by age 28-29 has a negative effect ($\beta = -.015$, $p < .05$) on earnings growth between the ages of 28-29 and 34-35 (full model not presented). Thus, while drug use has a positive relationship with earnings in the twenties, it yields a lower rate of wage growth that persists even after cessation of use.

Although at first paradoxical, this interpretation of a labor market effect is not inconsistent with the time-lagged positive effect of voluntary job changes on earnings observed over time (see Figure 1). In parallel to the labor market theory of wages, the job search paradigm proposes that certain job shifts early in a career represent attempts by young workers to improve the match between their own needs and aspirations and the characteristics and rewards of a job. We hypothesize that those most likely to follow this pattern are young workers who do not use drugs.

In a second step, re-estimation of the three-wave causal model separately for men who never used drugs (Figure 2) and those who did (Figure 3) support the hypothesis and provides further insights into the mechanisms that underlie the observed relationships between drug use and earnings. The two models are identical, allowing a test of the equality of paths among drug users and nonusers through the multisample analysis feature of LISREL.

The models provide several noteworthy findings.

FIGURE 2: Three-Wave Causal Model of Job Changes and Earnings: Nondrug Users[a]



$\chi^2 = 55.23$, df = 51, p = .318, AGFI = .878

[a] Male full-time workers at age 24-25, working at age 28-29 and 34-35. Standardized coefficients are shown. Race, year of schooling, self-reported health, and marital status are controlled for, but not shown. Nonsignificant paths are also not shown. (N = 161).

+ $p < .10$ * $p < .05$ ** $p < .01$

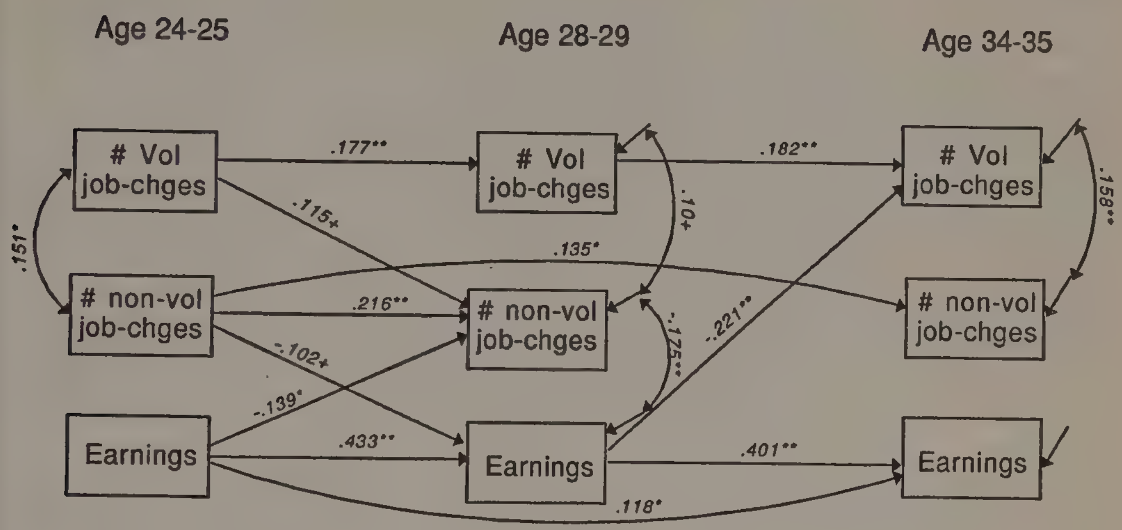
(1) For everyone, mobility early in one's career, i.e. by age 24-25, is crucial to long-term wage gains by age 34-35, either directly or indirectly. Mobility between ages 24-25 and 28-29 has no direct impact on wages at age 34-35.

(2) For all individuals, earnings in the mid-thirties are directly determined by earnings in the late twenties rather than directly by job mobility in the preceding six years.

(3) Nonvoluntary job changes are related to lower wages from the mid- to the late twenties for all individuals (see Figures 2 and 3). Drug users, however, experience more such types of job turnover than nonusers early in their career: $\bar{X} = 1.20$ for users, .68 for nonusers, difference significant at $p < .01$. There is a correlation of .225 ($p < .01$) between illicit drug use and nonvoluntary job leaves at age 24-25 (see Figure 1).

(4) Voluntary job mobility by the mid-twenties has a time-lagged positive effect on the wage growth of nonusers by the mid-thirties ($\beta = .161$, $p < .05$ for nonusers, −.060, nonsignificant, for users). However, a test of the equality of these two paths indicates that they are not statistically significantly different

FIGURE 3: Three-Wave Causal Model of Job Changes and Earnings: Ever Drug Users[a]



χ^2 = 39.57, df = 49, p = .830, AGFI = .938

[a] Male full-time workers at age 24-25, working at age 28-29 and 34-35. Standardized coefficients are shown. Race, year of schooling, self-reported health, and marital status are controlled for, but not shown. Nonsignificant paths are also not shown. (N = 240).

+ p < .10 * p < .05 ** p < .01

from each other. The relatively small size of the sample may be in part responsible for this result. It may still be the case that drug users do not reap the positive benefits of voluntary job leaves experienced by nonusers.

(5) The effect of earnings in the late twenties on earnings in the mid-thirties is smaller for users ($\beta = .401$, $p < .01$) than for nonusers ($\beta = .557$, $p < .01$). The difference between the two paths is statistically significant at $p = .01$. The higher wages of drug users early in their careers do not lead to higher wage growth subsequently.

Thus, drug users suffer three types of handicaps relative to nonusers: (a) They appear not to experience the time-lagged positive effects of voluntary job leaves on wage growth. (b) They experience stronger cumulative negative effects on wage growth from nonvoluntary job changes early in their career, from the mid to the late twenties, since they experience many more such leaves than nonusers. And, as noted above, earnings in the late twenties, and not job mobility, determine earnings in the mid-thirties. Finally, (c) drug users

Conclusion

Given the adverse impact of drug use, especially illicit drug use, on an individual's physical and psychological well being and the assumption that occupational achievement is associated with ability and productivity, drug use would be expected to lead to lower levels of productivity and consequently to lower labor force achievement and lower earnings. Most existing research reports the opposite with respect to earnings: the effect of illicit drug use is positive (Gill & Michaels 1992; Kaestner 1991, 1994; Kandel, Mossel & Kaestner 1987; Register & Williams 1992), or nonsignificant (Kandel & Davies 1990; Kaestner 1994). However, prior research on the impact of illicit drug use has been consistently based on samples of young adults, the majority of whom are in their twenties. The opportunity to observe individuals beyond their twenties over a longer period of labor force participation provides new insights regarding the impact of drug use on labor force performance, specifically earnings and rates of wage growth.

This study reveals that the impact of drug use on earnings depends upon the stage in the life-span. The controversial positive effect of illicit drug use on men's earnings disappears by the mid-thirties. Illicit drug use between the ages of 28-29 and 34-35 is related to lower earnings growth by age 34-35. Translated in terms of percentage changes, an additional year of using other illicit drugs between the ages of 24 to 29 leads to a 4% increase in rate of earnings growth but use between the ages of 29 to 35 leads to a 7% decrease in growth rate for each additional year of use. Mullahy and Sindelar's (1993) analysis of the impact of alcoholism on income found positive effects through the late twenties and negative effects at older ages. Furthermore, the authors also concluded that alcoholism was more highly related to labor force participation than to wages, a conclusion similar to the one we have reached with respect to illicit drug use. Economists stress that the full returns of higher educational attainment for earnings may not be achieved for many years after entry into the labor force. The point of overtaking may be as long as eight years (Griliches 1976, 1980; Mincer 1974). Similarly, the negative impact of drug use on earnings appears to take more than a decade to manifest itself and is observed among those who stopped using drugs as well as among those individuals who continue to use illicit drugs at a time in the life cycle when most of their peers have terminated their use.

Our prior results (Kandel & Yamaguchi 1987) led us to expect that the underlying process between lowered earnings and illicit drug use might be accounted for by increased job turnover. The relationships between drug use, job turnover and earnings are complex. The results of the causal analysis suggest that the effect of job turnover is indirect and is due to two different processes, both originating in the mid-twenties. At that time, drug users are more likely than nonusers to experience nonvoluntary rather than voluntary job changes. Nonvoluntary job turnover is related to lower subsequent growth in wages. In addition, drug users do not appear to experience the time-lagged positive effects of voluntary job mobility experienced early in their career on their wages later in their career. (Note also, that, controlling for prior job

mobility, the effect of drug use on job mobility appears only early in a work career, up to age 28-29. However, job mobility at that age does not directly predict reduced subsequent earnings by age 34-35.)

Thus, the causal model that we estimated also provides some important insights regarding processes related to labor force participation and psychosocial development in adulthood other than those specifically related to drug involvement. By differentiating type of job separation, we have documented that not all job mobility leads to lower income (Blank 1989; Sicherman & Galor 1990). Voluntary job changes eventually improve workers' wages among those who do not use drugs while nonvoluntary changes negatively affect wages for all workers. As mentioned, Topel and Ward (1992) associate early job changes with the movement toward mature careers. Our findings showing the different effects of early voluntary job mobility on earnings growth for drug users and nonusers suggests that individuals may differ for unobservable reasons in the types of voluntary changes they make, and, consequently, that drug use aids in identifying this individual heterogeneity.

One interpretation of the processes underlying the opposite relationships between drug use and earnings at different life stages is based on labor market theories of life-cycle compensation contracts and investment in training. We speculate that, early in their careers, drug users may choose higher paying jobs over higher subsequent growth potential, while nonusers make the opposite trade. These latter workers may accept low wages so as to obtain jobs that are ports of entry to job ladders that pay higher earnings in the future and because they have an incentive to share in the costs of firm-specific training. Our analyses have provided only a preliminary test of these hypotheses, which remain to be tested further.

Other interpretations can also be advanced. It may not be until the testing phase, possibly during a longer period of attachment to a particular work site, that the employee's wages would reflect levels of productivity. As suggested by French (1993), wages may not be a good indicator of an individual's loss of work productivity. An employer may not be able to adjust an individual's wages to the individual's specific level of productivity. Rather, increased marginal costs resulting from a drug user's lower productivity will be reflected in lower wages for all workers or higher prices paid by consumers. Or, the potential negative impact of drug use on productivity due to its negative physical effects may reflect lengthy cumulative use and may take many years to become manifest. Alternatively, the apparent effect of drug use may mask the effect of other unmeasured variables, such as personality or motivational variables. It is these variables, and their changes over time, which would account for the divergent relationships observed between income and drug use in the twenties and in the thirties.

A life-span perspective, however, specifies that the positive association between illicit drug use and earnings is typical of earlier as opposed to later stages in individuals' work careers, when the association turns negative. When the human capital-earnings relationship is viewed as a process whereby younger workers engage in training investments, job search, and long-term employment contracts to enhance future earnings potential, a predictable pattern of life-cycle earnings by drug-use status emerges. Training investments and

long-term employment contracts require that workers sacrifice present earnings for higher wage growth. If drug users are less willing to make these early sacrifices, they could conceivably earn higher wages at younger ages, other things equal. We propose that a labor market theory of life-cycle compensation contracts and investment in training and the differential impact that different types of job changes have in the short and the long run among users and nonusers provide a partial accounting of the processes underlying the observed relationships between drug use and earnings.

Notes

1. With a completion rate of 81% of those living, 1,325 persons were interviewed in 1980, at the mean age of 24.7. Four years later in 1984, 1,222 persons, comprising 92.5% of the 1980 samples (excluding 4 deceased), were reinterviewed at age 28-29. Of these, 1,160 were reinterviewed in 1990 at age 34-35, with a completion rate of 95.7% (excluding 10 deceased persons).

2. We considered and rejected the possibility of estimating the reciprocal effects of drug use and earnings by instrumental variable techniques, since the causal model we tested, as described in 2b indicated no effect from earnings to drug use.

3. The variables included in the probit models to calculate the correction term for sample selection bias into employment for the 1984 and 1988 regressions are listed in Appendix C together with the coefficients. Two adolescent variables measured in 1971 were included in the probit models but not the regressions. The minor delinquency index is a seven-item self-reported measure that assesses delinquent activities during the three months preceding the 1971 survey, including being sent out of class by a teacher or taking things of little or some value. Self-esteem is a six-item short form of the Rosenberg's self-esteem scale (Rosenberg 1965). The LIMDEP program was used to calculate the inverse Mill's ratio (λ) from probit models.

4. An alternative method to control for selection effects is to estimate an endogenous switching regression model (Maddala 1983), which would require separating the sample into users and nonusers and appending selectivity terms constructed from first-stage probit estimates of the drug-use choice to each earnings regression (Gill & Michaels 1992; Kaestner 1991). These probit estimates would be based, in large part, on those same variables believed to be related to the onset on drug use. Three factors caused us to reject this formulation. First, separating the sample into use categories would result in very small sample sizes for the earnings regressions. Second, our interest is in measuring the effect of a continuous measure of cumulative drug use on earnings. The standard switching regression formulation requires a discrete endogenous threshold variable. Finally, these models are very sensitive to deviations from normality. Thus, including those factors leading to drug use directly into the earnings regression is the preferred strategy for controlling for selection effects in the present case.

APPENDIX A: Means and Standard Deviations on Selected Variables among Males Who Worked in Full-Time Civilian Nonfarm Jobs, in 1984 and 1990

	1984			1990		
	Mean	Std. Dev.	Missing Percent	Mean	Std. Dev.	Missing Percent
Background						
Parents' highest schooling	13.13	2.75	2.5	13.22	2.75	2.7
Race (1 = black, 0 = nonblack)	.07	.25	0.0	.06	.23	0.0
1971 High school						
Peer activity index (10 to 40)	22.69	6.54	27.9	22.53	6.57	28.1
Peer activity missing[a]	.28	.44	0.0	.28	.45	0.0
Grade average (1 = F . . . 5 = A)	3.49	.78	0.0	3.49	.76	0.0
Individual charateristics (at time of interview)						
Married (1 = Yes)	.54	.50	0.0	.67	.47	0.0
Total years of schooling	13.94	2.25	0.4	14.17	2.40	0.0
Self-rating of health status (1-4)	3.32	.65	0.4	3.33	.63	0.6
Delinquency within last year	.49	.97	0.0	.37	.97	0.4
Work related characteristics (from 1971 up to interview)						
Years of work experience	8.02	2.35	0.0	13.01	2.50	0.2
Years in this job	3.68	2.70	0.0	5.92	4.42	1.9
Avg. hours worked per week	45.75	10.54	5.4	46.56	10.53	4.0
No. of voluntary job changes	1.52	1.45	0.0	2.07	1.83	0.0
No. of involuntary job changes	.88	1.59	0.0	1.12	1.87	0.0
No. of other job changes	.80	1.11	0.0	1.04	1.40	0.0
Log annual salary (in $1,000s)	3.12	.57	2.3	3.56	.53	1.0
Drug use (from 1971 up to interview in years)						
Cigarette smoking (1/2 pack a day)	3.55	4.40	0.0	5.29	6.42	0.0
Near-daily alcohol use	2.14	3.41	0.0	3.02	4.53	0.0
Monthly marijuana use	2.90	3.74	0.0	3.61	4.79	0.0
Monthly other illicit drugs use	1.71	3.25	0.0	2.18	4.15	0.0
Total N >		419			439	

[a] Missing because respondent was absentee in 1971 school survey.

APPENDIX B: Estimates from Multiple Regressions Predicting Log Annual Salary by Ages 28-29 (1984) and 34-35 (1990) among Male Workers with a Full-Time Civilian Nonfarm Job[a]

Covariates	By Age 28-29 (1984) r	Model 1[d]	Model 2[d]	By Age 34-35 (1990) r	Model 1[d]	Model 2[d]
Background						
Parents' highest schooling	.106*	.007	.016	.099+	-.002	.004
Race (1 = black, 0 = nonblack)	-.080+	-.140	-.210	-.170***	-.253**	-.023
1971 High school						
Peer activity index (10 to 40)[b]	.131**	.013**	.012*	.042	.005	.005
Grade average (1 = F . . . 5 = A)	.121**	-.002	-.007	.130**	-.040	-.059
Individual characteristics (at time of interview)						
Married (1 = Yes)	.164***	.155**	.099	.108*	.072	.036
Total years of schooling	.244***	.080***	.071***	.381***	.091***	.083***
Self-rating of health status (1-4)	.230***	.140***	.101*	.102*	.013	-.049
Delinquency within last year	.056	.020	.011	-.034	-.005	-.020
Work related characteristics (up to interview)[c]						
Years of work experience	-.066+	.131*	.146*	-.140**	.078	.063
Squared years work experience	-.048	-.005	-.006	-.147**	-.003	-.003
Years in current job	.043	.033	.031	.045	.022	.028+
Squared years in job	-.015	-.005+	-.004	.020	-.001	-.001
Avg. hours worked per week	.036	.003	.002	.225***	.008***	.008***
No. of voluntary job changes	-.059	.002	-.006	.052	.049***	.042**
No. of involuntary job changes	-.224***	-.069***	-.066***	-.164***	-.027*	-.022+
No. of other job changes	-.192***	-.106***	-.109***	-.073+	-.033*	-.027+
Drug use (cumulative years up to interview)						
Cigarette smoking (1/2 pack a day)	-.158**	-.014*	-.020*	-.136**	.001	-.002
Daily alcohol use	-.037	.001	-.004	-.078+	-.003	-.007
Monthly marijuana use	.037	.021**	.021*	-.102*	-.003	.003
Monthly other illicit drugs use	.040	-.001	-.001	-.030	.004	.006
Correction term (λ)	—	—	-.705*	—	—	-.872***
Constant		.367	.665		1.235*	1.965***
Adjusted R^2		.250	.258		.229	.259
Total N		419			439	

[a] Cumulative work and drug use included in models.
[b] Missing dummy for peer activity included in the model.
[c] Counted cumulatively since 1971.
[d] Model 1, uncorrected estimates; model 2, estimates with correction term for sample selection.

+ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

APPENDIX C: Factors Affecting the Probability of Full-Time Employment in 1984 (Age 28-29) and 1990 (Age 34-35) among Males: Probit Estimates

Covariates	1984	1990
Background		
Parents' highest schooling	-.036	.014
Family income other than salary[a]	.003	-.021**
Race (1 = black, 0 = nonblack)	.122	-.686*
1971 High school		
Peer activity index (10 to 40)	.018	-.004
Grade average (1 = F . . . 5 = A)	-.045	.046
Minor delinquency index[a]	-.133+	-.029
Missing in 1971	-.449*	-.035
Adult characteristics at 1984/ 1990		
Age	.083	.027
Married (1 = Yes)	.306+	.311+
Total years of schooling	.076	.026
No. of children in household[a]	.034	-.049
Self-rating of health status (1-4)	.009	.042+
Self-esteem[a]	.743**	.427**
Delinquency within last year	.066	.047
Drug use (between two interviews in years)		
Cigarette smoking (1/2 pack a day)	.040+	.018
Daily alcohol use	.026	.009
Monthly marijuana use	-.007	-.026
Monthly other illicit drugs use	-.001	.001
Constant	-4.256	-2.160
χ^2	26.680	36.598
Total N	539	538

[a] Variables included in probit but not the regression models in Table 2 and Appendix B.

+ $p < .10$ * $p < .05$ ** $p < .01$

References

Altonji, Joseph, and Robert Shakokto. 1985. *Do Wages Rise with Job Seniority?* Cambridge, Mass: National Bureau of Economic Research.

Bachman, Jerald G., Patrick M. O'Malley, and Lloyd D. Johnston. 1984. "Drug Use among Young Adults: The Impacts of Role Status and Social Environment." *Journal of Personality and Social Psychology* 47:629-45.

Blank, Emily C. 1989. "Mobility and Wage Growth of Male Household Heads." *Applied Economics* 21:475-85.

Blau, Peter M., and Otis Dudley Duncan. 1967. *The American Occupational Structure*. Free Press.

Borjas, George J. 1981. "Job Mobility and Earnings over the Life Cycle." *Industrial and Labor Relations Review* 34:365-76.

Carmichael, Lorne H. 1989. "Self-Enforcing Contracts, Shirking, and Life-Cycle Incentives." *The Journal of Economic Perspectives* 3:65-83.

Dolton, Peter J., and Gerald H. Makepeace. 1987. "Interpreting Sample Selection Effects." *Economics Letters* 24:373-79.

Featherman, David L., and Robert M. Hauser. 1978. *Opportunity and Change*. Academic Press.

Fligstein, Neil, and Wendy Wolf. 1978. "Sex Similarities in Occupational Status Attainment: Are the Results Due to the Restriction of the Sample to Employed Women?" *Social Science Research* 7:197-212.

French, Michael T. 1993. "The Effects of Alcohol and Illicit Drug Use in the Workplace: A Review." *Journal of Employee Assistance Research* 10:98-110.

Gill, Andrew M., and Robert J. Michaels. 1992. "Does Drug Use Lower Wages?" *Industrial and Labor Relations Review* 45:419-34.

Griliches, Z. 1976. "Wages of Very Young Men." *Journal of Political Economy* 84:69-85.

____. 1980. "Expectations, Realizations, and the Aging of Young Men." Pp. 1-21 in *Research in Labor Economics*, edited by R.G. Ehrenberg. JAI Press.

Griliches, Zvi, Bronwyn W. Hall, and Jerry A. Hausman. 1978. "Missing Data and Self-Selection in Large Panels." *Annales de l'INSEE* 30-31:137-76.

Griliches, Zvi, and William M. Mason. 1973. "Education, Income and Ability." Pp. 285-316 in *Structural Equation Models in the Social Sciences*, edited by Arthur S. Goldberger and Otis Dudley Duncan. Seminar Press.

Grossman, Mary. 1972. "On the Concept of Health Capital and the Demand for Health." *Journal of Political Economy* 80:223-55.

Hauser, Robert M., and T.N. Daymont. 1977. "Schooling, Ability and Earnings: Cross-Sectional Evidence 8 to 14 Years after High School Graduation." *Sociology of Education* 50:182-206.

Heckman, James J. 1980. "Sample Selection Bias as a Specification Error with an Application to the Estimation of Labor Supply Functions." Pp. 206-39 in *Female Labor Supply: Theory and Estimation*, edited by J.P. Smith. Princeton University Press.

Heckman, James J., and Solomon Polachek. 1974. "Empirical Evidence on the Functional Form of the Earnings-Schooling Relationship." *Journal of the American Statistical Association* 69:350-54.

Jencks, Christopher, Susan Bartlett, Mary Corcoran, James Crouse, David Eaglesfield, Gregory Jackson, Kent McClelland, Peter Mueser, Michael Olneck, Joseph Schwartz, Sherry Ward, and Jill Williams. 1979. *Who Gets Ahead? The Determinants of Economic Success in America*. Basic Books.

Jessor, Richard, John E. Donovan, and Frances M. Costa. 1991. *Beyond Adolescence. Problem Behavior and Young Adult Development*. Cambridge University Press.

Johnson, Robert J., and Cedric Herring. 1989. "Labor Market Participation among Young Adults. An Event History Analysis." *Youth & Society* 21:3-31.

Johnston, Lloyd D., Gerald Bachman, and Patrick M. O'Malley. 1984. *Monitoring the Future: Questionnaire Responses from the Nation's High School Seniors 1983*. Ann Arbor, Mich.: Institute of Social Research, University of Michigan, 1984.

Jöreskog, Karl G., and Dag Sörbom. 1989. *LISREL 7 Users's Reference Guide*. Mooresville, IN: Scientific Software.

Kaestner, Robert. 1991. "The Effect of Illicit Drug Use on the Wages of Young Adults." *Journal of Labor Economics* 9:381-412.

Kaestner, Robert. 1994. "New Estimates of the Effect of Marijuana and Cocaine Use on Wages." *Industrial and Labor Relations Review* 47:454-70.

Kandel, Denise B. 1984. "Marijuana Users in Young Adulthood." *Archives of General Psychiatry* 41:200-209.

Kandel, Denise B., and M. Davies. 1990. "Labor Force Experiences of a National Sample of Young Adult Men: The Role of Drug Involvement." *Youth and Society* 21:411-45.

Kandel, Denise B., P. Mossel, and R. Kaestner. 1987. "Drug Use, The Transition from School to Work and Occupational Achievement in the United States." *European Journal of Psychology of Education* 2:337-63.

Kandel, Denise B., V.H. Raveis, and J.A. Logan. 1983. "Sex Differences in the Characteristics of Members Lost to a Longitudinal Panel: A Speculative Research Note." *Public Opinion Quarterly* 47:568-76.

Kandel, Denise B., and K. Yamaguchi. 1987. "Job Mobility and Drug Use: An Event History Analysis." *American Journal of Sociology* 92:836-78.

Lazear, Edward P. 1979. "Why Is There Mandatory Retirement?" *Journal of Political Economy* 87:1261-84.

Maddala, G.S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Mare, Robert D., C. Winship, and W.N. Kubitschek. 1984. "The Transition from Youth to Adult: Understanding the Age Pattern of Employment." *American Journal of Sociology* 90:326-57.

Mensch, Barbara M., and Denise B. Kandel. 1988. "Dropping Out of High School and Drug Involvement." *Sociology of Education* 61:95-113.

Meyer, Robert H., and David A. Wise. 1982. "High School Preparation and Early Labor Force Experience." Pp. 277-339 in *The Youth Labor Market Problem: Its Nature, Causes, and Consequences*, edited by R.B. Freeman and D.A. Wise. University of Chicago Press.

Mijares, John C. 1992. "Labor Market Effects of Drug Use on Young Men." Unpublished paper. December

Miller, Judith D., Ira Cisin, and Adele Harrell. 1978. *Highlights from the National Survey on Drug Abuse: 1977*. Washington, D.C.: George Washington University Social Research Group.

Mincer, Jacob. 1974. *Schooling, Experience, and Earnings*. Columbia University Press.

Mincer, Jacob, and B. Javanovic. 1981. "Labor Mobility and Wages." Pp. 21-63 in *Studies in Labor Markets*, edited by S. Rosen. University of Chicago Press.

Mullahy, John, and Jody L. Sindelar. 1991. "Gender differences in labor market effects of alcoholism." *American Economic Review* 81 (Papers and Proceedings): 161-65.

Mullahy, John, and Jody L. Sindelar. 1993. "Alcoholism, Work, and Income." *Journal of Labor Economics* 11:494-520.

Newcomb, Michael D. 1988. *Drug Use in the Workplace*. Auburn House.

Newcomb, Michael D., and Peter M. Bentler. 1988. *Consequences of Adolescent Drug Use: Impact on the Lives of Young Adults*. Sage.

O'Donnell, John A., Harwin Voss, Richard R. Clayton, Gerald T. Slatin, and Robin G.W. Room. 1976. *Young Men and Drugs — A Nationwide Survey*. NIDA Research Monograph No. 5. Government Printing Office.

Register, Charles A., and Donald R. Williams. 1992. "Labor Market Effects of Marijuana and Cocaine Use among Young Men." *Industrial and Labor Relations Review* 45:435-51.

Robbins, C., Stevin S. Martin, and Howard B. Kaplan. 1985. "Adolescent Marijuana Use and the Transition to Adulthood: The Case of Work Role Attainment and Performance." Unpublished paper. January.

Rosenberg, Morris. 1965. *Society and the Adolescent Self-Image*. Princeton University Press.

Sicherman, Nachum, and Oded Galor. 1990. "A Theory of Career Mobility." *Journal of Political Economy* 98:169-92.

Topel, Robert H. 1991. "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority." *Journal of Political Economy* 99:145-76.

Topel, Robert H., and Michael P. Ward. 1992. "Job Mobility and the Careers of Young Men." *Quarterly Journal of Economics* 107:439-77.

U.S. Department of Labor, Employment and Training Administration. 1991. *Dictionary of Occupational Titles*. 4th ed. Government Printing Office.

White, Helene R., Angela Aidala, and Benjamin Zablocki. 1988. "A Longitudinal Investigation of Drug Use and Work Patterns among Middle-Class, White Adults." *The Journal of Applied Behavioral Science* 24:455-69.

Willis, Robert J. 1986. "Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions." Pp. 525-602 in *Handbook of Labor Economics*. Vol. 1, edited by O. Ashenfelter and R. Layard. Elsevier Science Publishers.

Yamaguchi, Kazuo, and Denise B. Kandel. 1985a. "On the Resolution of Role Incompatibility: Life Event History Analysis of Family Roles and Marijuana Use." *American Journal of Sociology* 90:1284-325.

____. 1985b. "Dynamic Relationships between Premarital Cohabitation and Illicit Drug Use: A Life Event History Analysis of Role Selection and Role Socialization." *American Sociological Review* 50:530-46.

Zarkin, Gary A., Michael T. French, and J. Valley Rachal. 1992. "The Relationship between Illicit Drug Use and Labor Supply." Unpublished paper. February

Pathways to Voluntarism: Family Socialization and Status Transmission Models*

THOMAS JANOSKI, *Duke University*
JOHN WILSON, *Duke University*

Abstract

Participation in voluntary associations is usually explained by a Weberian theory that uses human capital variables; however, Durkheimian theory suggests the importance of parental socialization and family status variables. Using a three-wave panel study, this article models the changes in social participation that people experience while moving from high school to parenthood. Voluntary participation is accounted for in part by the transmission of socioeconomic status, but family socialization through example and value modeling are often more important. When self-oriented (occupation and profession) and community-oriented (service, church, community, fraternal, and neighborhood) types of participation are distinguished, the status transmission theory explains self-oriented but not community-oriented participation. Family socialization explains community-oriented but not self-oriented participation. Thus, theories of voluntarism must be differentiated according to the type of voluntary association concerned.

People join voluntary associations to achieve a variety of personal objectives which they could not obtain by acting alone. Participation in voluntary associations can cover such disparate activities as "team bowling, card-playing, voting as a member of a committee, belonging to a neighborhood improvement association, protesting racial discrimination, interacting with kin and friends, [and] contributing to the Community Chest" (Edwards & Booth 1973:1). In some of these cases the organization resembles a "mutual benefit association" because the prime beneficiary of the collective action is the membership itself. In other cases, the organization resembles a "service organization," because a client group (e.g., the homeless), not the membership, is the prime beneficiary (Blau & Scott 1962:140). The question thus arises whether the same sociological theory can be used to explain participation in both types of organization. In this article,

** We wish to thank Lynn Nonnemaker for research assistance. We benefited from critical readings by Charles Clotfelter, Anne Tickamyer, and the anonymous reviewers.* *Direct correspondence to Thomas Janoski, Sociology Department, Box 90088, 268 Sociology/Psychology Building, Duke University, Durham, NC 27708. FAX: 919-660-5623.*

we propose a conceptualization of voluntary association behavior in which these activities are divided into self-oriented associations (e.g., professional and union groups) and community-oriented associations (e.g., community volunteers, church-related organizations). We argue that different explanations are needed to resolve why people engage in each type of voluntary association, and that the factors that explain participation in one type have little to do with participation in the other.

Theory

Two sets of theories have guided the study of voluntary association memberships. The first is associated with Emile Durkheim. One of the theories within this set Hechter (1987:20) calls the "normativist perspective," because it sees joining as an expression of the solidarity that comes from adherence to a set of corporate obligations. The obligations, or norms, are learned in the same way that other norms are learned, informally through family and friends, and formally through schools, churches, and the workplace. It is also possible that such norms are acquired from the mass media (Payne et al. 1972:228). Of all agents of socialization, the most influential is the family (Gecas 1981:170). If this is true, and if the normativist perspective has anything to contribute to the understanding of voluntarism, the disposition to join should be transmitted directly from one generation to another. Mechanisms of family socialization include modeling, reinforcement, role playing, labeling, and social comparisons (Bandura 1977).

As Hechter (1987) points out, the normativist perspective does not exhaust the possibilities of Durkheimian theory. From the writings of Durkheim, and a contemporary, Georg Simmel, has emerged the "structuralist perspective." From this perspective, whether or not people will join and become active in secondary associations depends very much on the network of social relations in which they are embedded.[1] Voluntary action is contingent on the degree of social integration of the individual into society. As noted by Moen et al. (1989), social integration is a concept historically applied to collectivities, but it can also describe the multiple roles of individuals, where it "denotes the concrete involvement of individuals with various aspects of a collectivity" (635). This aspect of Durkheimian theory would lead us to expect that employed people would be more active than the unemployed, that married people are more active than single, that parents are more active than the childless, and so on. In other words, people become more or less active depending on the number of social ties they have.

This aspect of Durkheimian theory has a number of obvious applications: it suggests that participation in the labor force promotes voluntarism; that people with roots in the community will be more active; that joining produces more joining. It also suggests that families influence voluntarism in ways other than those suggested by the normativist perspective. From that perspective, voluntarism is the expression of internalized norms inculcated by parents. Children resemble their parents because of what they have been taught by them. From the structuralist perspective, voluntarism is the expression of structural isomor-

phism: children resemble their parents because their roles and statuses become identical; i.e., they have themselves become parents (besides, in most cases, remaining children).

Only a few studies of voluntary associations have included family variables in the analysis (for notable exceptions see Klobus-Edwards et al. 1984; Knoke & Thompson 1977). This neglect is surprising because a number of studies have shown that both family of origin and family of procreation are important influences on two very common types of voluntarism — political party membership and church participation (Jennings & Niemi 1981; Sandomirsky & Wilson 1990). These studies suggest that socially active parents produce socially active children. This would provide support for the social learning theory that people "get most of their ideas by observing the actions of others who serve as models" (Heiss 1981:102). The family of procreation effect is explained by structural theory: when people become like their parents in having children they become active in social groups like their parents.

The second set of theories of voluntary association memberships is associated with Max Weber. A Weberian theory of voluntary association memberships attributes more significance to socioeconomic factors: voluntary action is driven, or made possible, by socioeconomic interests and resources, and will change as these interests and resources change. Thus, when McPherson (1981:718) finds that people in higher status occupations are typically more active in voluntary associations, he attributes this to their being more attractive to, and more likely to be recruited by, a voluntary organization, and to their being more likely to reap "social contact" rewards from voluntarism. It is indeed true that, for many people in high-status occupations, voluntary association membership is almost part of the job. In many cases membership dues are paid by the employer. Not surprisingly, active membership in voluntary associations is almost always found to be positively correlated with income, educational and occupational status levels (Axelrod 1973:40; Babchuk & Booth 1973:26; Pearce 1993:65). While these separate dimensions of socioeconomic status tend to be quite highly correlated, there appears to be some agreement that education is the best predictor of social participation among this group of variables (McPherson & Lockwood 1980).

Socioeconomic status, of course, can be traced in part to family background, but the principal determinant of voluntarism according to Weberian theory is *current* socioeconomic resources and interests, since life chances are affected by market position. The theory does not, therefore, rule out the influence of the family, but assumes that it will be indirect, through the station in life parents provide their children. Describing the pattern of intergenerational transmission of values, Dalton (1982) contrasts the "attitudinal" pathway (141), which would depend for its efficacy on direct personal contact, presumably before the child leaves the home, and the "social milieu" pathway, which consists of the kind of social relationships a given parental socioeconomic background makes possible. Acock (1984) supports the view that parent-child similarities are as likely to be the product of the intergenerational transfer of status as of individual psychosocial influence. Glass, Bengtson and Dunham (1986) also suggest that generational similarities and differences can be due to either socialization or transmission of status.

This article's purpose is to show that one does not have to choose between these sets of theories to explain social participation. Voluntary action can take a wide variety of forms, and it is possible that one sociological theory explains one form better than another. It is a standard feature of sociological writings on voluntary associations to categorize them into different types. The distinction between "instrumental" and "expressive" is perhaps the most familiar (Edwards & Booth 1973), but the parallel distinction between "purposive" and "solidary" associations is also often drawn (Clark & Wilson 1961). Instrumental or purposive groups are those aimed mainly at some tangible outcome for the participating individuals or a client group, while expressive or solidary groups are those organizations where participation itself is the primary source of satisfaction (e.g., fellowship). If there are indeed different types of voluntary associations, then perhaps there are different reasons for belonging to them and different social factors associated with membership.

Most scholars of voluntary associations agree on the difficulty of categorizing them (Pearce 1993:18). Not only are many organizations multipurpose but, on the individual level, membership can mean different things to different members. Even those who have adopted the instrumental/expressive distinction are obliged to admit mixed "instrumental-expressive" types into their analyses (e.g., Edwards & Booth 1973:1-2). The instrumental-expressive dichotomy itself has been criticized for being gender biased, since most of the organizations typically categorized as instrumental are heavily male consigning females to the expressive category. Just as important, the instrumental/expressive categorization leaves little room for organizations whose primary purpose is building community. Organizations are assumed to be either intended to further the interests of their members or purely inward-looking and expressive.

Community building is an important aspect of voluntarism, and a conceptual scheme that obscures this function is deficient. Our method of categorizing voluntary associations is to think of them in terms of self- versus community-orientation. Some forms of social participation have as their chief goal furthering an individual's self-interest, while others are intended more to protect the interests of a community as a whole. Our self-oriented type is close to the instrumental type, but our community-oriented type is intended to cluster together those associations whose ostensible mission is the improvement of a community. This distinction is similar to that proposed by Pearce (1993) who, drawing on Thompson (1976), distinguishes between voluntary associations and voluntary organizations. Voluntary associations are groups of people who share an interest and have agreed to pursue it jointly. "Association members are both 'owners' and 'consumers' of the association's output and need please no one but themselves" (Pearce 1993:18). This corresponds to our self-oriented type. Voluntary organizations, on the other hand are engaged in generalized exchanges with nonmembers. "They produce a service or product for outsiders" (Pearce 1993:18). This corresponds to our community-oriented type.

Most social surveys of voluntary association memberships use a fairly standard coding scheme now associated with the General Social Survey. It prompts respondents to indicate membership in any of the following kinds of groups: church-connected, lodges or fraternal, neighborhood clubs, sports teams, informal clubs like sewing circles, business or professional groups, service

groups, ethnic or racial groups, and labor unions. We define business/professional, trade union and veterans groups as self-oriented. We define church, fraternal, neighborhood, and service organizations as community-oriented. We thus try to distinguish voluntary activity which is primarily oriented to increasing or protecting benefits, furthering business or occupational interests and the like, from those activities that are oriented primarily to caring, community welfare, and social improvement. The division is, of course, simplistic. It ignores the fact that many organizations are multipurpose, or provide side benefits in addition to their main benefits. It also ignores differences within each category: business groups vary in their degree of emphasis on service to the community, just as churches and fraternal orders vary in the degree to which they cater mainly to the needs of their own members.

Some organizations exist for companionship and entertainment. These groups, such as sports and hobby organizations, neither promote income and occupational interests nor benefit the larger community in any direct way. Consequently, we eliminated sports teams, informal clubs, and ethnic and racial associations from both the self- and community-oriented categories. Much volunteer work, of course, centers around youth-oriented programs like Little League Baseball. Participation in these kinds of activities is, however, often difficult to classify. Sports activities that one engages in for amusement hardly qualify as voluntarism. Making the distinction is even more difficult in the case of ethnic and racial associations. These groups often provide dinners, dances, and other entertainment, but they also involve both community action for the improvement of members' life changes and opportunities for community building with other groups. Because these groups had a strong entertainment element and involved significant mixing of self and community elements, we eliminated them from the analysis rather than force them into one or the other category.[2]

It is our contention that Weberian theory accounts for participation in self-oriented organizations, while Durkheimian theory accounts for participation in community-oriented organizations. Participation in self-oriented organizations is based largely on a calculus of costs and benefits, or structured largely by selection procedures that favor certain socioeconomic groups over others because the exchange of membership for service is beneficial to the organization. Conversely, participation in community-oriented groups is only marginally dependent on resources and is more of an expression of the individual's socialization and degree of social integration. If our argument is sound, it also follows that Durkheimian theory cannot account for participation in self-oriented organizations, and Weberian theory cannot account for participation in community-oriented organizations.

Data Source

The two theories could be tested using cross-sectional data, since one indicates that voluntary association memberships are associated with current degree of social integration while the other sees voluntarism as associated with current social status. However, they are best tested by panel data, for they actually

describe different paths to voluntarism. In the one path, parents bestow status on their off-spring, which in turn leads to voluntarism, and they have this effect regardless of their own degree of participation or their children's degree of social integration. In the other path, parents teach lessons in voluntarism and provide role models for their children, and they have this effect regardless of the human capital they bestow on their children.

This study uses data from the Youth-Parent Socialization Panel Study (YPSP), collected by the Survey Research Center at the University of Michigan (Jennings & Niemi 1981). The first wave of the study in 1965 yielded 1,699 randomly selected high school seniors from a national probability sample of high schools, representing a 99% response rate. A randomly selected parent of each student was also interviewed, resulting in 1,562 completed interviews for a 93% response rate. The second wave was collected in 1973, and retained 1,348 (80.8%) of the youths from the original panel, and 1,179 of the parents (75.5%). In 1982, 1,135 of the students were re-interviewed and 898 of the parents. For the purposes of this study, we concentrated on that group of students who were interviewed in all three waves and who had a parent interviewed in both 1965 and 1973 ($n = 924$).

Dependent Variables

Our dependent variables describe levels of participation in voluntary associations. Participants in the Youth-Parent Socialization Panel study were shown a list of organizations "to which people may belong" and asked to indicate which they were members of, if any. Those who indicated they belonged to an organization were asked separately to specify their level of activity in the organization, ranging from "not very active," through "fairly active," to "very active." Types of organization on the list were: church-connected, fraternal, neighborhood, sports, informal, professional or business, service, ethnic, and labor unions. In supplementary questions, respondents were given two chances to mention organizations they belonged to that were not on the provided list and were asked to indicate their level of activism in them.

We recoded these basic variables to make them more manageable and to facilitate a better test of our theory. First, we "folded in" the answers to the supplementary questions, recoding the answers to allocate them to existing categories. We added veterans' groups to the trade union category; political clubs and youth-related groups to the neighborhood category; charitable enterprises to the service group category; and country clubs and hobby groups to the informal club category (not included in the final analysis). The numbers involved in this re-coding were small. Second, we combined the answers to the first two questions to create a variable measuring membership and level of activism simultaneously. The recoding reads: 0 = not a member; 1 = member, but inactive; 2 = member, fairly active; 3 = member, very active. Third, we created a participation index for each respondent in each year. We did this by summing the scores 0-3 for each of the organizations listed.

Our fourth step was to select two groups of organizations to be our measure of self and community-oriented participation. In the first category we

placed business/professional, trade union, and veterans groups. In the second category we placed church-related, fraternal, neighborhood (and youth related and political), and service organizations. The result of these efforts is two groups of dependent variables: self-oriented measures of participation in 1973 and 1982; community-oriented measures of participation in 1973 and 1982. These categories measure very different types of social participation. There is no correlation between membership in self-oriented and community-oriented organizations among our respondents in 1973. The correlation between self-oriented and community-oriented memberships in 1982 is significant ($p = .0013$) but somewhat small ($r = .10$).

Independent Variables for the Socialization Theory

To predict community-oriented voluntarism in 1982, our study will use parent's voluntarism in the early waves of the panel study as a measure of role-modeling. The socialization theory predicts that parental example will have an immediate impact on the child. We therefore examine the student's own social participation while in high school. In addition, since it is reasonable to suppose the voluntary association memberships are somewhat "sticky," we also use the student's social participation in adolescence to predict later voluntarism. The data set does not contain a direct measure of voluntary association memberships for the student while in high school. As a surrogate, we created a measure of student's high school social participation from a number of questions put to the high school seniors concerning the level of their participation (if any) in a number of extracurricular activities (Hanks 1981). These included being a member of a school athletic team, school band, school debating team, publication board, hobby club, school subject club, occupation club, and neighborhood club. Students were also asked if they had ever run for political office, and we included this in our measure of social participation. Responses were coded to match adult social participation, from 0 for not a member to 3 for a very active member. The student social participation index for 1965 was created by summing the scores on these variables. We made no attempt to discriminate at this stage between self- and community-oriented participation.

The same measures of voluntary association memberships are used for the parent as specified in the section on the dependent variable, with one exception. In the first wave of the survey, there are three cases: both parents were interviewed, only the mother was interviewed, or only the father was interviewed. Using one parent's data as a basis for the independent variables would have resulted in a large number of missing cases (mothers are missing cases where the father is interviewed, fathers are missing cases where the mother is interviewed). We therefore decided to create a single "parent variable." This consists of the mother's score where the mother was interviewed and the father's score where the father was interviewed. In order to remain consistent with our parents' socioeconomic measure, where the father's score is used in the case of both parents being interviewed, we decided to take the father's voluntarism score where scores from both spouses are available. Since mothers

might be said to have a slightly stronger tendency to volunteer than fathers, this is a conservative measure of parental influence on volunteering.

Thus, there are six participation variables for parents: (1) the overall participation index for 1965 and 1973, (2) the community-oriented index for 1965 and 1973, and (3) the self-oriented index for 1965 and 1973. In the analysis that follows, we use only variables in item 2.

The assumption of family responsibilities is an important part of the Durkheimian model. We measured both marital and parental status. A marital status question was asked of students in 1973 and again in 1982. We collapsed the responses into married (coded 1) and single (coded 0). The parental status variable came from different questions in 1973 and 1982. Respondents were asked in 1973 to list the members of their household and their relationship to them. A measure of number of children was computed from this variable. In 1982, respondents were asked if they had any children and, if so, how many were sons and how many were daughters. These responses were combined to yield a count of number of children. Marital status and parental status are, of course, quite highly correlated. Multicollinearity problems makes using both variables in a regression model troublesome. Since our theory places most weight on the role of children in fostering voluntary association memberships, we use the parental variable in the regression analyses.[3]

Independent Variables for the Status Transmission Theory

Human capital is typically measured in terms of income, education, and occupational status. The Youth-Parent Socialization Study data set provides information on all three for both parents and children. The income variable was the simple amount of income coded into increments of $2,000 until "$35,000 and above." For education, students in 1973 and 1982, and parents in 1965 and 1973, were asked their current level of education. The item used in this study codes responses according to the number of years completed. Noncollege postsecondary education received an additional one year added to high school education. The occupational coding scheme varied from year to year: we converted all measures of occupation into the Duncan SES code for 1970. The code runs from 1 to 99. As previous researchers have discovered, the three measures of socioeconomic status tend to be highly correlated with each other and with the dependent variable. In the multivariate analysis, therefore, we dropped occupational status from the model.

Although coded in the same way as the students, parent's income and occupation presented special problems given our need to construct a single-parent variable by combining the father's and mother's responses into a single variable. Education of the mother may be comparable to the father, but many of the women were housewives in both 1965 and 1973 and their reported income and occupational status could not be considered an accurate record of a family's socioeconomic circumstances. To deal with this problem, we took the mother's report of head of household's income in 1965 and the mother's report of spouse's income in 1973 in order to construct a head-of-household variable. The result is a variable that largely reflects father's income and occupation;

however, mother's income and occupation are used in single-parent households or when the mother is the head of household.

Men are more likely to belong to voluntary associations than women (Edwards & Booth 1973:23). Since we are interested in the effects of status transmission, and given the likelihood that men and women would be differently affected by these status transmission processes, we include gender as a control in the status transmission model. For consistency, we introduced gender into the socialization model, but it was unrelated to participation in 1965, volunteering in 1973, or volunteering in 1982. Since it made no difference to the other associations in this model, we dropped it from the socialization analysis. Gender was coded 1 for male and 2 for female.

Methods

Our analysis is guided by two hypotheses:

Hypothesis 1. Community-oriented participation but not self-oriented participation is explained by socialization variables.

Hypothesis 2. Self-oriented participation but not community-oriented participation is explained by status-transmission variables.

To test these hypotheses, we constructed two models, both aimed at predicting variation in the dependent variable, the degree of participation in voluntary associations in 1973 and 1982. The Weberian model assumes that social status will have an impact on current social participation and that current social status is to some extent dependent on parent's social status. The Durkheimian model assumes that social participation is partly inherited directly from parents' own participation and is also a function of current family statuses. The main part of the analysis consists of using Durkheimian theory to estimate variations in community-oriented participation and Weberian theory to estimate variations in self-oriented participation. We then look at how well each theory explains variation in the "wrong" kind of voluntarism (e.g., socialization model explaining self-oriented voluntarism).

Results

Table 1 reports rates of activism for students in three waves and parents in two. Allowing respondents multiple mentions meant that ranges on these variables varied from 0-13 to 0-25. The means were therefore standardized to make comparisons possible. The 1965 "voluntarism" variable for students is not directly comparable to the others in the table since it measures participation in extracurricular activities while in high school and not voluntary associations. Nevertheless, it is a decent proxy measure of the individual's disposition to social participation.

In a recent cross-national comparison of voluntary association memberships, using data gathered at the same time as the third wave of the Youth-Parent Socialization Panel Study, Curtis et al. (1992) found that 72.7% of Americans belong to some kind of voluntary association. This was easily the highest

TABLE 1: Standardized Means and Unstandardized Ranges for Participation Variables[a]

	Total Voluntary Association Participation	Community-Oriented VA Participation	Self-Oriented VA Participation
Student at mean age 18			
Standardized participation index for 1965	1.6136	NA	NA
Unstandardized range	0 to 17	NA	NA
Student at mean age 26			
Standardized participation index for 1973	.8971	.5386	0.5920
Unstandardized range	0 to 13	0 to 6	0 to 6
Student at mean age 35			
Standardized participation index for 1982	1.1204	.7710	.7314
Unstandardized range	0 to 21	0 to 12	0 to 6
Parent at mean age 45			
Standardized participation index for 1965	.9046	.7565	.5461
Unstandardized range	0 to 25	0 to 12	0 to 6
Parent at mean age 55			
Standardized particpation index for 1973	1.1994	.7923	.5266
Unstandardized range	0 to 13	0 to 6	0 to 6

[a] NA Indicates that the student's high school participation could not be divided into community-oriented and self-oriented variables in 1965.

percentage among the nations compared, and much higher than the mean reported in Table 1 for the nationally representative sample used in this study. This discrepancy can be accounted for in a number of ways. First, the wording of the question used in the study drawn upon by Curtis et al. (1992) is ambiguous as far as religious organizations are concerned, for the interviewer prompt includes "churches and religious organizations." Many respondents are clearly including church membership in this list, which is inflating the rate. We do not include church memberships (e.g., Catholic Church) in our list, but membership in church-related organizations (e.g., Knights of Columbus, Legion of Mary). Second, the data reported in Table 1 measure not only membership but activism. Curtis et al. also recognize that membership is not the same as

participation and report "working membership" rates: the percentage for active members falls to 31.8% for all associations and 18.9% for all secular organizations, participation rates much closer to but still lower than those reported from the Youth-Parent Socialization Study.

Where most respondents are reporting no active voluntary association memberships at all, it is to be expected that the overall means will be low. The column for total voluntary association participation shows high school students have the highest rate. This is largely an artifact of the different question used. High school extracurricular activities are far more structured and integrated into the school social system than are voluntary associations in the wider community. In addition, some high school extracurricular activities are, if not mandatory, subject to peer group pressure to belong.

Voluntary association memberships in early adulthood have a mean about half that of high school social participation. This is partly an artifact of the stricter measure of voluntarism used in 1973, but it is also in line with a pattern of voluntary association memberships over the life course, in which memberships are low among young adults (Curtis et al. 1992:147; Knoke & Thompson 1977).

Student participation increases from age 26 to 35. Parents' participation also increases from age 45 to 55. The parents' rate at age 45 is slightly lower than the student's rate at age 35. This means an artificial "rolling cohort effect" is not produced by the table. On the whole, however, previous findings that voluntary association participation increases with age are replicated. The break in the upward curve between students who are 35 in 1982 and parents who are 45 in 1965 has more to do with the differential cohort effects of parents growing up from 1938 to 1946, when voluntary associations had been significantly reduced by the depression and state mobilization, and their sons and daughters growing up from 1965 to 1973, when voluntary associations greatly increased due to increased social mobility and antiwar activity.

In comparing columns 2 and 3, a life cycle and an age effect are seen with the two different types of organizations. At age 26, students are more involved in self-oriented associations, but by the time they reach 35 students are slightly more involved in community-oriented participation. The parents are much more involved in community-oriented participation at the age of 45, and this gap grows even larger by the time they reach the age of 55. Thus, over time, community-oriented participation increases, while self-oriented activity declines after a high point at the age of 35. We interpret this as young adults needing to make a successful transition into the labor force. In this process, many join and participate in self-oriented voluntary associations to promote their career. However, by the age of 35 their careers tend to be more or less established. These adults now have families and their voluntary association activities shift to family concerns which involve a more community-oriented kind of participation. Both age and life cycle effects are important and complementary.

The zero-order correlation coefficients between the independent variables and the three measures of voluntary association memberships are shown in Table 2. The first columns show the correlations for overall participation; the second column shows the correlations for community-oriented participation; and the third column shows the correlations for self-oriented participation. The

TABLE 2: The Relationship Between Total, Community-Oriented, and Self-Oriented Voluntary Association (VA) Participation in 1982 and Two Sets of Independent Variables[a]

	Total Voluntary Association Participation, 1982	Community-Oriented VA Participation, 1982	Self-Oriented VA Participation, 1982
Socialization model			
Student's VA participation			
Community-oriented in 1973	.15 (.0001)**	.25 (.0001)**	-.01 (.7157)
Extracurricular in 1965	.05 (.1529)	.08 (.0333)**	-.03 (.3512)
Parent's VA participation			
Community-oriented in 1973	.11 (.0038)**	.14 (.0001)**	-.02 (.5380)
Community-oriented in 1965	.06 (.0836)	.03 (.4601)	.03 (.3505)
Student's family status			
Marital status in 1982	.06 (.0870)	.17 (.0001)**	-.03 (.3766)
Number of children in 1982	.01 (.9626)	.14 (.0002)**	-.08 (.0203)#
Marital status in 1973	-.04 (.2082)	.08 (.0184)**	-.12 (.0006)##
Number of children in 1973	-.08 (.0266)##	.01 (.8726)	-.09 (.0137)##
Status transmission model			
Student's self-oriented VA participation in 1973	.21 (.0001)**	.06 (.1100)	.33 (.0001)**
Student's SES and gender			
Income in 1982	.20 (.0001)**	-.01 (.7960)	.36 (.0001)**
Education in 1982	.21 (.0001)**	.08 (.0181)**	.24 (.0001)**
Occupational status in 1982	.13 (.0009)**	.06 (.1546)	.13 (.0007)**
Income in 1973	.14 (.0004)**	.04 (.2524)	.19 (.0001)**
Education in 1973	.13 (.0003)**	.04 (.3068)	.17 (.0001)**
Occupational status in 1973	.08 (.0295)*	.02 (.6014)	.10 (.0115)*
Gender	-.16 (.0001)**	-.01 (.7877)	-.25 (.0001)**
Parent's SES			
Income in 1973	.08 (.0254)*	.01 (.6973)	.07 (.0495)**
Education in 1973	.07 (.0386)*	.01 (.9768)	.10 (.0060)**
Occupational status in 1973	.07 (.0521)	.01 (.7551)	.06 (.0937)
Income in 1965	.08 (.0414)*	.01 (.7441)	.01 (.7360)
Education in 1965	.07 (.0429)*	.01 (.9879)	.10 (.0070)**
Occupational status in 1965	.08 (.0245)**	.01 (.9605)	.07 (.0395)**

[a] Zero-order correlation coefficients with significance levels in parentheses

* $p < .05$ ** $p < .01$
# $p < .05$, but opposite hypothesized direction
$p < .01$, but opposite hypothesized direction

top panel displays the coefficients for the socialization model; the bottom panel shows the coefficients for the status transmission model. For overall student voluntary association participation in 1982, few socialization variables are significant, while the effects of many of the socioeconomic status variables are both strong and significant. As might be expected, students' own socioeconomic status is more strongly associated with voluntary association memberships than is their parent's socioeconomic status.

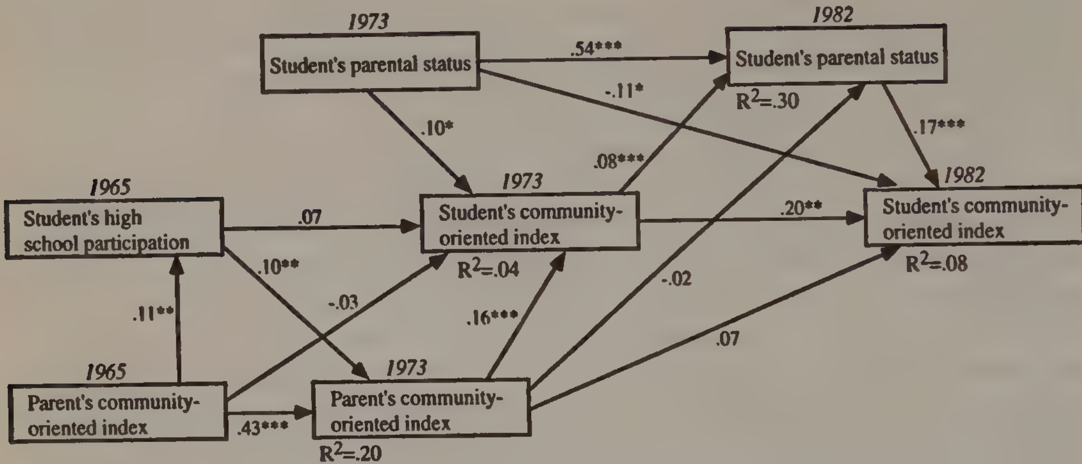
The main point of Table 2 is to contrast columns 2 and 3. In the case of column 2, we hypothesized that most of the variables in the top panel would be significant while few of the variables in the bottom panel would be so. This is indeed true. All but number of children in 1973 and parent's voluntary association memberships are significant in the top panel; only education in 1982 is significant in the lower panel. In the case of column 3, we anticipated that no variables would be correlated in the top panel and all would be correlated in the bottom panel. The top panel actually displays more significant variables than we had anticipated. However, the significant variables are negatively, not positively, related to membership in self-oriented associations. Married people with children are less likely to belong to this type of organization. The bottom panel fulfills our expectations. Only parent's income in 1965 and parent's occupational status in 1973, factors quite far removed from the dependent variable, are unrelated. Female gender is negatively related to membership in self-oriented associations.

The zero-order coefficients provide preliminary but nevertheless substantial support for our two major hypotheses. However, they say little about how the several variables are related to each other and to the dependent variable. In other words, they say little about the different "paths" to voluntary action earlier described. For this, path analysis is necessary.

Figure 1 presents the socialization model used to estimate variations in community-oriented participation in 1982. Students whose parents were active in community-oriented associations in 1965 are more likely to be active themselves while in high school. However, high rates of social participation in high school have no carry-over effect on voluntarism once the student leaves school. There is an indirect path of influence of parents on children, however, and that is through parent's community-oriented activism in 1973 which is, in turn, dependent on their prior activism in 1965. Parents who are active in 1973 are more likely to have children who are also active. This association holds, regardless of the student's prior social participation, and regardless of their parental status, which is itself positively related to community-oriented activism.

The fact that parental influence is sustained across the two waves is especially noteworthy. It is to be expected that parents will have more impact on their children when those children are in their mid-twenties than when those children are in their teens. The gap between the parents in 1965, with an average age of 45 and the youths, with an average age of 18, is larger in subjective terms than the gap in 1973, when the parents are 55 and the students are 26. It is undoubtedly true, also, that the social participation of parents and children becomes more similar over this life-course transition because, structur-

FIGURE 1: Socialization Path Model Explaining Community-Oriented Voluntary Association Participation



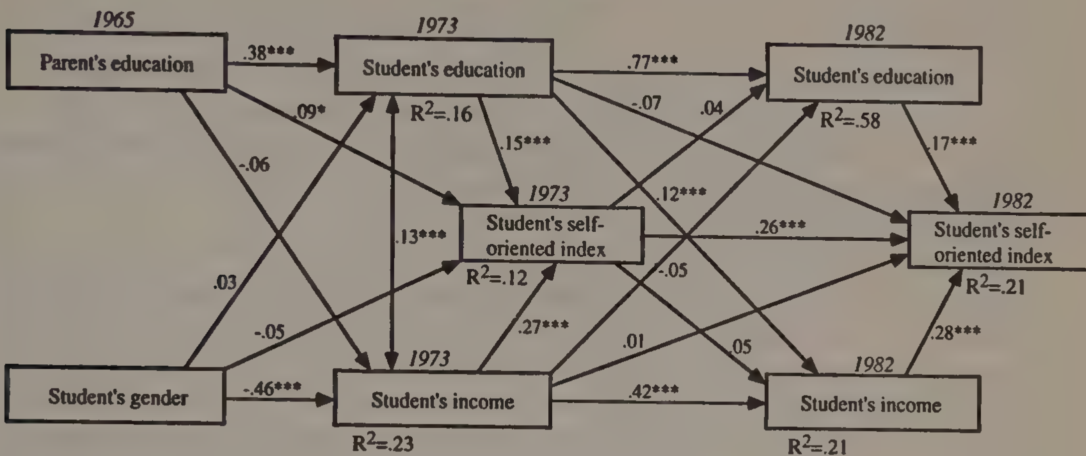
* $p < .05$ ** $p < .01$ *** $p < .001$

ally, the parents and children are becoming more similar; but this effect of parents on children is net of parental status for the children.

Community-oriented activism among students who have reached their mid-thirties, as expected, is influenced by their prior voluntary association memberships. However, the relationship is not particularly strong. This might be explained by the data reported in Table 1 showing that many respondents became active between the ages of 26 and 35. The impact of parent's participation has by this time faded. As predicted, parental status has a positive effect on voluntary association memberships in 1982, slightly stronger than the effect in 1973. The negative effect of parental status in 1973 on community volunteering in 1982 is interpretable in light of the parental status in the 1982 variable. Those who had become parents by 1973 had begun having children as early as nineteen. Their children could have been anywhere between age 1 and 7 in 1973. By 1982 these children would have been between 10 and 16. Those who became parents sometime between 1973 and 1982, however, would have had much younger children in 1982. We believe that the age of the children helps explain the negative sign. People who had not added to the number of children by 1982 had, by that time, children who were too old to pull their parents into social participation. The path to community-oriented voluntarism is thus clear from this diagram, where the effects of parental modeling and social integration are visible, despite the fact that not much of the total variance in community-oriented voluntary association memberships is explained.

Figure 2 reports the results of the test of the self-oriented model. This model accounts for much more of the variance (20%) than the socialization model does for community-oriented participation. Family status background, as measured by the level of education of the parent, has a small but significant impact on the likelihood of volunteering in 1973, even when a number of other factors are

FIGURE 2: Status Transmission Path Model Explaining Self-Oriented Voluntary Association Participation



* p < .05 ** p < .01 *** p < .001

controlled. Family background also has an indirect effect, as predicted by the theory, through the respondent's own educational achievement by 1973. Highly educated parents produce highly educated children who, in turn, are more likely to be active in voluntary associations of the self-oriented type. Both education and income have positive effects on voluntary association memberships in 1973. Indeed, the effect of income is stronger than that of education. Gender has no direct effect on the chances of voluntarism. However, women earned less money than men in 1973 and this in turn makes it less likely that they will be active.

Like community-oriented participation, self-oriented participation reproduces itself. People who were active in the second wave are more likely to be active in the third wave than those who were inactive. However, both education and income have powerful and highly significant effects on 1982 membership in this type of voluntary association. Again, level of income is more important than level of education, suggesting that membership, and particularly activism, is based partly on the ability to pay (e.g., professional dues) or make donations.[4]

Figures 1 and 2 provide substantial support for the theory that voluntary association memberships are acquired through two different pathways, depending on the type of membership involved. Community-oriented activism is almost directly inherited: the activity and example of the parents has a powerful influence on that of the offspring. When those offspring begin to resemble their parents in roles and responsibilities, they begin to act like them. Membership in self-oriented voluntary associations, on the other hand, is inherited from parents indirectly. Parents make it possible for children to acquire the human capital that makes membership in such organizations desirable and feasible, although they are more likely to do this for their sons than their daughters.

The argument that socialization variables explain community-oriented voluntarism and human capital variables explain self-oriented voluntarism would be more convincing if it could be shown that socialization and role modeling have no impact on self-oriented participation and that the inheritance of human capital has no impact on community-oriented participation. In other words, if it could be shown that the "wrong" theories do not work. Table 2 has already strongly suggested that this is the case. Zero-order correlations between status variables and community-oriented participation are weak at best; those between socialization variables and self-oriented participation are either insignificant or negative.

The most appropriate test of the "wrong" theory would be to repeat the path analyses reported in Figures 1 and 2, switching the independent variables. Since (as we anticipated from the correlation table) this resulted in a path diagram with a mass of insignificant paths, we have conducted a simpler contrast, reported in Tables 3 and 4, where we show only direct effects on voluntarism in 1973 and 1982. The tables report the results of separate regression analyses, estimating models using the "correct" variables (columns 1 and 2) and models using the "incorrect" variables (columns 3 and 4).

Table 3 describes the results of using socialization variables to estimate variation in community-oriented and self-oriented participation. Unstandardized coefficients are reported, unlike the path diagram, to make comparison across models possible. Our expectation is that the socialization model should do a better job of explaining columns 1 and 2 than columns 3 and 4. As reported in the path diagram in Figure 1, community-oriented participation in 1973 is chiefly accounted for by parent's community-oriented participation and by the respondent's parental status. There are no other direct effects. This is confirmed in column 1. Column 2 confirms that parental status in 1982 has an effect on voluntarism in 1982, net of the rate of voluntarism in 1973. Column 3, using the "wrong" variables to estimate variations in self-oriented participation in 1973 shows that parent's social participation has no effect and parental status is negatively related. Column 4 shows no direct effect of socialization variables on self-oriented participation, except for a very small (and uninterpretable) direct effect of parent's participation in 1973. The variance explained by the correct model (in column 2) is about four times that explained by the incorrect model (in column 4) using the unadjusted R^2s.

Table 4 describes the results of using status transmission variables to estimate variation in self-oriented and community-oriented participation. Columns 1 and 2 merely repeat the results of the path diagram, reporting unstandardized coefficients. They confirm the importance of education and income in both waves. The indirect effect of parent's education on 1973 voluntarism is not, of course, revealed in this table. Columns 3 and 4 demonstrate that status transmission variables are totally incapable of accounting for differences in community-oriented participation. The variance explained drops from 20% in column 2 to 1% in column 4. The correct model explains fifteen times as much variance as the incorrect model.

TABLE 3: Socialization Variables Explaining the Community-Oriented and Self-Oriented Models of Voluntary Association (VA) Participation: Multivariate Regression Equations[a]

Independent variables	Dependent Variables (1) Community-Oriented Participation, 1973	(2) Community-Oriented Participation, 1982	(3) Self-Oriented Participation, 1973	(4) Self-Oriented Participation, 1982
Parent's participation in community-oriented VAs in 1965	-.010 (.538) [.017]	-.027 (.240) [.052]	.001 (.925) [.015]	.0492* (.039) [.021]
Student's participation in high school extracurricular activities in 1965	-.022 (.073) [.012]	.008 (.592) [.016]	.016 (.121) [.011]	-.005 (.729) [.014]
Parent's participation in community-oriented VAs in 1973	.103** (.004) [.029]	.077* (.046) [.038]	.006 (.781) [.024]	-.083 (.015) [.034]
Student's parental status in 1973	.123* (.014) [.050]	-.186* (.019) [.079]	-.104* (.018) [.044]	-.033 (.640) [.071]
Student's participation in community-oriented VAs in 1973		.268** (.001) [.054]		.009 (.846) [.049]
Student's parental status in 1982		.207** (.001) [.062]		-.066 (.846) [.049]
R^2	.041	.078	.015	.019
Adjusted R^2	.034	.068	.008	.009
(N = 584)				

[a] Unstandardized betas, significance in parentheses, and standard errors in brackets.

* $p < .05$ ** $p < .01$

TABLE 4: Status Transmission Variables Explaining the Self-Oriented and Community-Oriented Models of Voluntary Association (VA) Participation: Multivariate Regression Equations[a]

	Dependent Variables			
	(1) Self-Oriented Participation, 1973	(2) Self-Oriented Participation, 1982	(3) Community-Oriented Participation, 1973	(4) Community-Oriented Participation, 1982
Independent variables				
Student's gender (1 = female)	-.043 (.625) [.086]	-.092 (.479) [.129]	-.012 (.914) [.109]	-.006 (.968) [.169]
Parent's education in 1965	-.031** (.009) [.012]	.015 (.324) [.015]	-.007 (.641) [.015]	-.012 (.561) [.020]
Student's education in 1973	.086** (.001) [.024]	-.052 (.259) [.046]	.004 (.882) [.030]	-.017 (.777) [.060]
Student's income in 1973	.068** (.001) [.011]	-.006 (.656) [.015]	-.025 (.070) [.014]	.036 (.065) [.019]
Student's participation in self-oriented VAs in 1973		.376** (.001) [.054]		.064 (.362) [.069]
Student's income in 1982		.068** (.001) [.012]		-.027 (.087) [.016]
Student's education in 1982		.089* (.015) [.037]		.052 (.276) [.048]
R^2	.117	.205	.007	.014
Adjusted R^2	.111	.195	.001	.003
(N = 595)				

[a] Unstandardized betas, significance in parentheses, and standard errors in brackets.

* $p < .05$ ** $p < .01$

Conclusion

The study of voluntary association memberships has always had to deal with the problem that not all voluntary associations are alike. Previously, voluntary associations have been categorized as either expressive or instrumental, a distinction with its roots in the structural functionalism of Talcott Parsons. While it is undeniable that social organizations can either pursue goals that are primarily intended for members of the organization or are intended for the broader community (of which members are a part), we have argued that a more accurate rendition of this distinction would be self- versus community-oriented, where the first group of organizations seeks benefits and rewards for its members directly, while for the second group, any personal direct benefits members receive are incidental to the main mission of the organization, which is to meet a perceived community need.

Implicit in the debate over types of voluntary associations is the assumption that they represent different forms of social participation and are motivated by different concerns. A natural corollary of this, but one that has not been explored previously, is that each type of voluntary action has different roots. This is the major argument presented in this article. Human capital theories of voluntary association memberships are valid, but only with respect to the self-oriented type of voluntary association. In short, the theory works best at explaining precisely the kind of "voluntarism" where some benefit can be promised and calculated. This kind of voluntarism has familial roots, but they lie in the kind of human capital people inherit from their parents. Membership in this type of organization is a form of social capital more likely to be possessed (and passed on) by the wealthy and highly educated.

Human capital theories do a poor job, however, of explaining membership in groups where direct pay-offs cannot be calculated. Where the organization's rewards go to the larger community, voluntarism takes on a different character. Here, no amount of human capital inherited from parents seems to make a difference. This does not mean that parents are without influence. Rather, they socialize by example and even, as far as we know, by direct contact and recruitment to the group in question. These kinds of associations are much more endogamous, passed on from family to family. This finding is not surprising given that religious organizations are included, but note that these are not church affiliations, or preferences, but active memberships in church-related organizations.[5]

The data also confirm how important social integration is to certain types of voluntarism. Thus, membership in community-oriented organizations is quite strongly related to having children — an expected and natural finding. Having children is negatively related, however, to membership in self-oriented groups. No doubt the ties that bind are different in each case. In the case of community-oriented participation, social integration means ties to neighborhood, schools, parks, and similar places of collective consumption, principally through children. If this is self-interest, it is self-interest once removed, through one's children, and it is self-interest broadly defined, where the goods received are public and not private. In the case of self-oriented participation, the ties that bind are probably more narrowly conceived. They are principally occupational

and professional ties. The ties are those of production rather than consumption. If the first group of organizations is seeking to improve the circumstances of consumption (e.g., housing, streets, libraries, health care), the second group is seeking to improve the circumstances of production (e.g., pension benefits, safety rules, business contacts).

Distinguishing voluntary associations along the lines of self- versus community-oriented organizations also prompts a re-thinking of gender differences in voluntarism. It is widely reported that men have higher rates of social participation than women net of human capital variables such as education (Knoke & Thompson 1977). It has also been surmised that men are more likely to be active in "instrumental" associations and women more active in "expressive" associations. The assumption seems to be that men's traditional gender roles assign them a more public and women a more private social life. However, we found no gender differences in rate of participation in community-oriented voluntary associations. Women have as high a rate of participation in community-oriented associations as men. This suggests that women care as much about the public household as the private, that they have as public a role as men, albeit of a different kind, a role born of their frequent role as mothers. While they might be dependent in the role of the wife, in the role of the mother, they are all the things men claim to be — active, productive, strong, empathetic, and altruistic (Rossi & Rossi 1990). Just as noteworthy, we found no direct effects of gender on membership in self-oriented voluntary associations. Rather, women's lower rate of participation in this kind of organization is a result of their lower educational and income levels. Women do not participate as much as men because they lack the human capital and are perhaps not seen as desirable as men in the role of volunteer or board member.

Our results suggest that further research needs to be done on the age and life-cycle effects of community- and self-oriented participation. The table we present has a cohort gap between students and parents. More direct longitudinal evidence on community and self-oriented participation over the life course on the same sample would be more convincing evidence on this point. In effect, the table suggests that people move from liberal principles of self-interest at an early age when they need to establish a career to more communitarian principles of community orientation at later ages when they are building a community for their children. They may be more clearly altruistic when they continue to volunteer for community-oriented organizations after their children grow up and move away.

Notes

1. Hechter (1987:25) is correct in identifying this structuralist aspect to Durkheim's theory, linking it to the work of Simmel, and distinguishing it from the normativist perspective. However, his own characterization of it is highly individualistic: "structuralists usually consider solidarity to be the product of rational individual action" (25). In a footnote he dismisses the structuralists' focus on social networks because "it is not apparent how it can account either for the genesis or the transformation of these structures." We reject this argument and make the case that social integration is a concept of much value in the study of voluntary associations.

2. We recognize that a person may altruistically belong to a self-oriented voluntary association (e.g., an accountant from a CPA's association volunteering to do the books for the Red Cross). Also, a person may have selfish reasons for belonging to a community-oriented association (e.g., a person whose business involves producing braille books may belong to the Lions, who benefit blind persons). However, on balance, these are exceptions to the rule. Our explanation is not totally deterministic, but rather statistical. We further justify our division of organizations by the fact that the self- and community-oriented indexes are poorly correlated with each other, and that entirely different independent variables explain the two.

3. We also constructed a family status variable by combining responses to the marital status and parental status questions. This variable was coded one if the respondent was married with children and zero if the respondent was married without children, had children but was not married, or was single and childless. Using this variable rather than the parental status variable did not significantly alter the results of our analyses.

4. We give causal interpretations to variables such as educational level and parental status measured simultaneously since neither is "instantly" achieved. It is just conceivable, of course, that voluntarism is the independent variable in both cases. People who get involved in community-oriented organizations are more likely to become parents. People who join self-oriented organizations thereby boost their income (the connection to education is less plausible). These counter-interpretations cannot be ruled out. However, they are less convincing than the interpretations given here, and they ignore both indirect effects and the fact that prior voluntarism is being controlled for in each case. That is, both types of voluntarism are the result of an accumulation of effects, so that simultaneous variables do represent prior events.

5. Studies of intergenerational transmission of values have frequently assumed that parents have a more certain and lasting impact on their children if the relationship between them is close. However, Smith (1982) found that degree of affection between parents and offspring has little to do with how much they share educational goals. The Youth-Parent Socialization Study contains a question on how emotionally close to their parents students feel, in each of the three waves. Using this item, we created a variable to capture the interaction between parental example and closeness. But there was no relationship.

References

Acock, Alan. 1984. "Parents and Their Children: The Study of Inter-Generational Influence." *Sociology and Social Research*. 68:151-71.

Axelrod, Morris 1973 "Urban Structure and Social Participation." Pp. 39-44 in *Social Participation in an Urban Society*, edited by John Edwards and Alan Booth. Schenkman Publishing.

Babchuk, Nicholas, and Alan Booth 1973 "Voluntary Association Membership: A Longitudinal Analysis." Pp. 23-37 in *Social Participation in an Urban Society*, edited by John Edwards and Alan Booth. Schenkman Publishing.

Bandura, Albert. 1977. *Social Learning Theory*. Prentice Hall.

Blau, Peter and W. Richard Scott. 1962. *Formal Organizations*. Chandler.

Clark, Peter, and James Q. Wilson. 1961. "Incentive Systems: A Theory of Organizations." *Administrative Science Quarterly* 6:129-66.

Curtis, James, Edward Grabb, and Douglas Baer. 1992 "Voluntary Association Memberships in Fifteen Countries: A Comparative Analysis." *American Sociological Review*. 57:139-52.

Dalton, Russell. 1982. "The Pathways of Political Socialization." *American Politics Quarterly* 10:139-57.

Edwards, John, and Alan Booth 1973 "Introduction to Social Participation." Pp. 1-18 in *Social Participation in an Urban Society*, edited by John Edwards and Alan Booth. Schenkman Publishing.

Gecas, Victor. 1981. "Contexts of Socialization." Pp. 165-99 in *Social Psychology,* edited by Morris Rosenberg and Ralph Turner. Basic Books.

Glass, Jennifer, Vern Bengtson, and Charlotte Dunham. 1986. "Attitude Similarity in Three-Generation Families: Socialization, Status Inheritance or Reciprocal Influence?" *American Sociological Review* 51:685-98.

Hanks, Michael. 1981. "Youth, Voluntary Associations and Political Socialization." *Social Forces* 60:211-23.

Hechter, Michael. 1987. *Principles of Group Solidarity.* University of California Press.

Heiss, Jerold. 1981. "Social Roles." Pp. 94-132 in *Social Psychology,* edited by Morris Rosenberg and Ralph Turner. Basic Books.

Jennings, M. Kent, and Richard Niemi. 1981. *Generations and Politics: A Panel Study of Young Adults and Their Parents.* Princeton University Press.

Klobus-Edwards, Patricia, John Edwards, and Ann Dewitt Watts. 1984. "Women, Work and Social Participation." *Journal of Voluntary Action Research* 13:7-22.

Knoke, David, and Randall Thompson. 1977. "Voluntary Association Memberships and the Family Life Cycle." *Social Forces* 56:48-65.

McPherson, J. Miller. 1981. "A Dynamic Model of Voluntary Affiliation" *Social Forces* 59:705-28.

McPherson, J.M., and W.G. Lockwood. 1980. "The Longitudinal Study of Voluntary Association Membership — A Multivariate Analysis." *Journal of Voluntary Action Research* 9:61-79.

Moen, Phyllis, Donna Dempster-McClain, and Robin Williams Jr. 1989. "Social Integration and Longevity." *American Journal of Sociology* 54:635-48.

Payne, Raymond, Barbara Pittard Payne, and Richard Reddy. 1972. "Social Background and Role Determinants of Individual Participation in Organized Voluntary Action." Pp. 207-50 in *Voluntary Action Research,* edited by David Smith, Richard Reddy, and Brian Baldwin. Lexington Books.

Pearce, Jone. 1993. *Volunteers: The Organizational Behavior of Unpaid Workers.* Routledge.

Rossi, Alice, and Peter Rossi. 1990. *Of Human Bonding.* Aldine de Gruyter.

Sandomirsky, Sharon, and John Wilson. 1990. "Processes of Disaffiliation: Religious Mobility Among Men and Women." *Social Forces* 68:1211-29.

Smith, David. 1982. "The Case for Parental Transmission of Educational Goals." *Journal of Marriage and the Family* 44:661-74.

Thompson, Victor. 1976. *Bureaucracy and the Modern World.* Morristown, New Jersey: General Learning Press.

The Impact of Alcohol, Divorce, and Unemployment on Suicide: A Multilevel Analysis*

THOR NORSTRÖM, *Stockholm University*

Abstract

Sociologists have taken little interest in alcohol abuse as a possible antecedent of suicide. However, a theoretical link between the two phenomena can be postulated that is consistent with Durkheim's suicide theory. In this article, various data and methods are triangulated to scrutinize the relationship between alcohol and suicide. The control variables include, among others, divorce and unemployment. The empirical basis comprises ecological and time-series data for Sweden. Special efforts are made to integrate the findings with results from existent individual-level studies. According to the findings, a good third of the male suicides are attributable to the alcohol factor. The impact of unemployment is also found to be fairly strong, and here the indirect effects seem to be at least as important as the direct ones. No relationship was found between divorce and suicide.

The psychiatric and the sociological literature on suicide differ conspicuously in the role attributed to alcohol abuse as an antecedent of suicide. Hence, a large number of psychiatric and epidemiological studies single out alcohol abuse as an important risk factor in suicide. In contrast, the alcohol factor is not generally considered in sociological studies on suicide. For example, the word alcohol is never mentioned in Stack's (1982) extensive review of the sociological literature on suicide. The neglect of the alcohol factor is further indicated by the fact that out of a total of 3,610 references on suicide listed in *Social Sciences Citation Index* 1981-93, only 28 mention alcohol.

A possible explanation of this state of affairs is that Durkheim in his influential theory of suicide explicitly dismissed alcohol as a causal factor of suicide (Durkheim [1897] 1966:77ff). In a critical review of Durkheim's theory, Skog (1991) shows that this view is not unequivocally supported by Durkheim's own data. For instance, a closer examination of the data pertaining to the French departments reveals a substantial correlation between alcohol and suicide, contrary to Durkheim's inferences from these data.

I wish to thank Eero Carrol, Ole-Jørgen Skog, Rolf Sundberg, and two anonymous referees for advice and comments. Direct correspondence to Thor Norström, Swedish Institute for Social Research, Stockholm University, S-106 91 Stockholm, Sweden.

Many sociologists may find it natural that alcohol abuse, as well as its causes and consequences, are a concern only to psychiatry and related disciplines. It is certainly true that personality and biological factors create a broad range of individual variation in the risk of developing a drinking pattern that we would label alcohol abuse. However, if our focus is rather on changes in the prevalence of alcohol abuse in society, then a sociological perspective is much more fruitful. An influential theorization of this kind is the network approach to explaining alcohol abuse which has been developed by Skog (1985a). The point of departure of this approach is the fact that drinking behavior is social behavior; individual drinkers influence each other's drinking habits (although not necessarily symmetrically). This influence is not only mediated directly through face-to-face interactions, as observed experimentally in actual drinking situations (Bruun 1959; DeRicco & Garlington 1978), but also indirectly. Individuals' drinking behavior is affected by individuals in their social networks who, in turn, are affected by other individuals. As a matter of fact, large-scale changes in consumption need not depend solely on external factors, such as prices; they may also be produced by microlevel fluctuations that are synchronized by the operation of long-range indirect ties (Skog 1986). Thus there is a significant cultural dimension to the dynamics of alcohol consumption; much empirical evidence shows that various consumption categories, including heavy drinkers, move up and down the consumption scale more or less in concert (Skog 1985a, 1985b). A further implication of this is the existence of a close relationship between per capita consumption and the prevalence of heavy drinking and alcohol abuse (Skog 1985a). These phenomena may therefore be conceived as social facts in Durkheim's sense of the word.

Just as a sociological perspective is relevant for understanding the genesis of alcohol abuse, so can its consequences (including suicide risk) be approached sociologically. More specifically, it seems possible to postulate a theoretical link between alcohol abuse and suicide which is consistent with Durkheim's ([1897] 1966) well-known assertion that the suicide rate is inversely related to the degree of social integration. If by social integration we are here referring to the extent and quality of interpersonal relations, there is much to suggest that alcohol abuse has disintegrating effects. Some plausible mechanisms will be outlined here.

A meta-analysis of some thirty experimental studies reveals that consumption of alcohol tends to elicit aggressive behavior (Bushman & Cooper 1990). Situations that engender frustration seem to increase the likelihood of this response (Gustafson 1986). The aggressive behavior induced by alcohol may come in various forms and degrees, although most of the existent evidence concerns physical violence. There is thus a high incidence (typically above 50%) of intoxication among violent offenders (see e.g., Mayfield 1976; Roslund & Larson 1979). The relationship between alcohol and violence is also borne out at the aggregate level; a temporal relationship between alcohol consumption per capita and various rates of criminal violence has been established for several countries (Lenke 1990; Skog & Bjørk 1978). There is also documentation of less dramatic consequences of alcohol that are detrimental for interpersonal relations. On the basis of a population survey, Hauge and Irgens-Jensen (1986)

thus report a positive correlation between alcohol intake and the incidence of reckless behavior.

What has been noted above are alcohol-induced behaviors on the part of the heavy drinker that tend to dilute his or her social network. In addition, we may assume a hostile or disapproving attitude from society at large toward alcohol abuse which may worsen the prospects of heavy drinkers to become socially integrated. One piece of evidence in support of this notion is the following finding from a Swedish survey study (Kühlhorn & Ramstedt 1993). Out of a list including various minority and deviant groups (inter alia alcoholics, immigrants and criminals), the respondent was asked which of these persons he or she would not accept as a neighbor. It appeared that "alcoholics" was the least accepted category (indicated by 44%, followed by "criminals," 35%). Similar results are reported in a Danish study (Gundelach 1992).

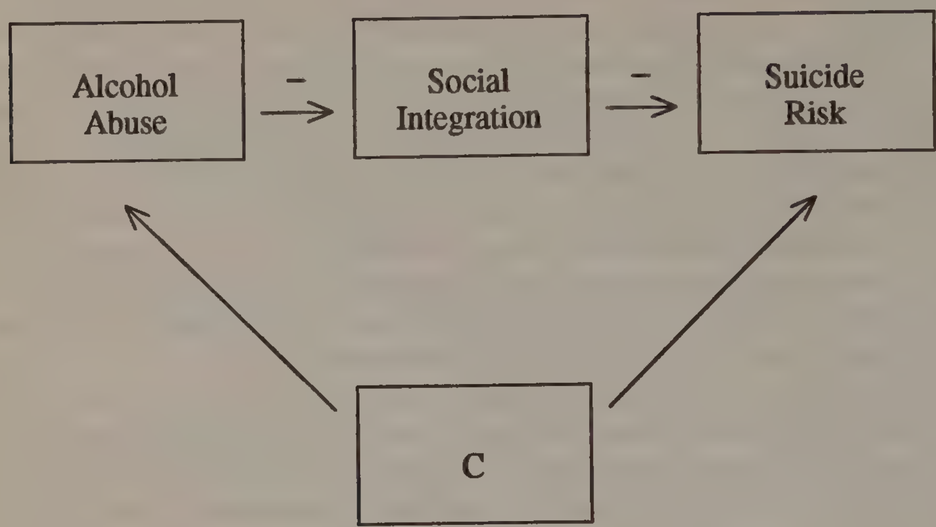
There are thus several reasons for believing that alcoholics and excessive drinkers are poorly integrated, having deficient social networks, which is also documented in the literature (Singer, Blane & Kasschau 1964; Straus & McCarthy 1951). It must be recognized that this may in part be the outcome of self-selection, in the sense that alcoholics have inferior capability for social interaction prior to the onset of alcoholism. However, a study by Philips (1976) indeed suggests that the progress of alcoholization is accompanied by a deterioration of social ties. In other words, the degree of estrangement increases for each stage of the alcoholization process. Two additional studies (Cahalan 1970; Cahalan & Room 1974) that explicitly address the temporal ordering between excessive drinking and complications with social relations also show that the former generally precedes the latter.

More direct evidence on the impact of excessive drinking on social integration and the suicide risk is found in Murphy's work (1992) which consists of detailed case studies of 50 consecutive alcoholic suicides in St. Louis (1968-71). A great majority (72%) of the subjects had little or no social support; relationships with spouses, children and friends were strained or terminated, and this was typically a result of the subjects' alcohol abuse. The fatality of this attrition of social support is indicated by the finding that one fourth (26%) of the alcoholic suicides had experienced an interpersonal loss in the last six weeks of their lives. This was "a last link to others /that/ had failed, generally as a consequence of the alcoholic's heedless disregard for its value to him or her." (Murphy 1992: 250.) The corresponding figure in an earlier study (Murphy & Robins 1967) is 32%, where it is also reported that such interpersonal loss was much less frequent among nonalcoholic suicides (3%).

The main feature of the above discussion of the link between alcohol and suicide is graphically summarized in Figure 1: Alcohol abuse tends to induce aggressive, reckless behavior that increases the risk of losing personal ties, i.e., leads to weaker social integration. This in turn, following Durkheim, increases the risk of suicide. To recognize the possibility of confounding influences, in Figure 1 C represents factors that are common causes of alcohol abuse and suicide risk, such as personality disorders.

The hypothesized association between alcohol and suicide is substantiated by a large number of epidemiological and psychiatric findings. Studies of individuals who commit suicide have repeatedly found a high incidence of

FIGURE 1: Hypothetical Model of the Link between Alcohol and Suicide



alcoholism (Beskow 1979; Robins et al. 1959). Further, several follow-up studies of alcoholics report grossly elevated risk-ratios for suicide (Dahlgren 1951; Nørvig & Nielsen 1956; Ritson 1977; Sundby 1967).

Since the prevalence of alcohol abuse is closely related to the overall level of drinking in the society (Skog 1985a) we should also expect an aggregate relationship between alcohol consumption and suicide. This is borne out in a growing number of studies. In an analysis of U.S. time-series data, Wasserman (1989) found that per capita alcohol consumption had a significant impact on the suicide rate. Also for Hungary (Skog & Elekes 1993), Norway (Rossow 1993), and Sweden (Norström 1988a) the effect of alcohol on the suicide rate is statistically significant and of approximately the same magnitude, while for Denmark the alcohol effect seems to be much weaker and detectable only in a time period with appreciable variation in consumption (Skog 1993; cf. Norström 1988a). A negative finding, with no significant alcohol effect, is reported for Finland (Norström 1988a).

Aim and Approach

Although the evidence presented above strongly suggests an association between alcohol and suicide, both the micro- and macrolevel studies conducted thus far have limitations. Findings from microlevel studies are compromised by the possibility of selection effects: alcoholics who commit suicide may already be at elevated risk prior to their drinking careers. The macrolevel findings, on the other hand, are always open to the objection of omitted variable bias. In this article I will triangulate various data and methods in order to scrutinize more

carefully the relationship between alcohol abuse and suicide risk. A number of other possible causes of suicide are included as control variables, which also makes it possible to assess the relative importance of the alcohol factor. The empirical basis of the study consists of ecological and time-series data for Sweden; in addition cross-level comparisons are made with existing individual-level studies. Since these studies pertain to men, the aggregate analyses are focused on the male suicide rate.

The rationale of triangulation is that a relationship that is supported by various kinds of data, analyzed by different methods, is more credible than a relationship that has been replicated several times, but by one method only. In the latter case there is always the risk that the finding is method-bound. When applying multiple testing under varied conditions, on the other hand, it is recognized that each test has its own weaknesses but that it is less likely that all tests are impaired by one common bias.

As already noted the major limitation of individual-level findings is the possibility of drift, or selection effects (Lieberson 1985). An essential comparative advantage of an aggregate effect estimate (whether obtained from time-series data or ecological data) is that it is expected to express the exogenous impact of the risk factor, net of selection effects (Norström 1988b). The rationale for this assertion is that the temporal variation of the risk factor (the effect of which is measured by the regression coefficient in a time-series analysis) should be due to changes in exposure and not to changes in self-selection. Analytically, this can be shown as follows. Consider the following micro model:

$$Y_{it} = C_i + \beta X_{it} + \epsilon_{it} \tag{1}$$

where Y_{it} is the response of individual i on the outcome variable at time t; C_i is the sum total of those unmeasured time-invariant attributes that are correlated with the outcome; and ε_{it} is a random error term. The potential bias in the estimation of β is due to the correlation between C_i and X_{it} that may arise because of self-selection. Now consider equation 1 in aggregate form:

$$Y_t = C + \beta X_t + \epsilon_t \tag{2}$$

As we can see the aggregation turns C_i into a constant which by definition is uncorrelated with X_t, and thus the source of bias is eliminated. To illustrate, let Y represent suicide, C personality disorders, and X unemployment. If personality disorders increase the individual's suicide risk as well as his risk of becoming unemployed, the estimate of β in the micro equation 1 will be biased upwards due to the selection effect. If we are willing to make the plausible assumption that variations in the prevalence of personality disorders at the aggregate level does not affect the unemployment rate, the estimate of β in the macro equation 2 is void of selection effects. (If this assumption is invalid, the estimate of β will be compromised by omitted variable bias.)

Thus both micro- and macrolevel studies have their weaknesses, but of different kinds. A micro finding that is replicated with macro data therefore gains in credibility from the fact that is not likely to have been produced by one common source of bias. However, strict comparisons of effect estimates from micro and macro studies are hampered by the use of different measures of association, typically relative risks and regression coefficients respectively. This calls for an impact measure that can be applied across different kinds of data,

and that can also serve our second purpose, namely to assess the relative importance of alcohol for suicide as compared to that of other causes. In research of the present kind, where the basic aim is to estimate a risk factor's effect on a specific form of mortality, such a standard is provided by an epidemiological measure: the attributable fraction (*AF*). It is usually calculated from micro data on the basis of two parameters: the relative risk (*RR*) and the fraction of the population (θ) which is exposed to the risk factor (e.g., Lilienfeld & Lilienfeld 1981:217):

$$AF = \frac{\theta(RR - 1)}{\theta(RR - 1) + 1} \qquad (3)$$

where the relative risk is defined as:

$$RR = \frac{\textit{Mortality rate for the exposed category}}{\textit{Mortality rate for the unexposed category}} \qquad (4)$$

The attributable fraction expresses the relative contribution of a risk factor to a specific cause of death — in other words, $AF * 100$ expresses how many percent mortality would be expected to decrease if the risk factor were eliminated.

The attributable fraction and the relative risk can be calculated on the basis of macro data as well (see Norström 1988b, for derivation of formulas). Consider a general macro model of the form:

$$Y = \alpha + \beta X + \epsilon \qquad (5)$$

where β is the unstandardized regression coefficient, *X* is the predictor (the risk factor), and *Y* is the outcome. Assume that *Y* is the mortality rate. The mortality rate produced by *X* will then be equal to $\beta * X$, and thus:

$$AF = \frac{\beta X}{Y} \qquad (6)$$

We may compute the attributable fraction for each unit of observation (i.e, each year if annual data are used) or for the average of *X* and *Y*. The measure is meaningful only if β is positive. The attributable fraction computed in this way has, theoretically, no upper limit. However, if a value exceeding 1 (or some other value that is implausible when compared with external information) is obtained, the estimate of β is evidently biased.

Given a semilogarithmic model (which will be applied in the present analysis) of the form:

$$Ln(Y) = \alpha + \beta X + \epsilon \qquad (7)$$

the formula[1] for the attributable fraction is:

$$AF = 1 - e^{-\beta X} \qquad (8)$$

To sum up, because different measures of association are so awkward to interrelate, cross-level comparisons seldom go beyond the sign and statistical significance of a relationship. However, by converting micro- and macrolevel estimates into a common impact measure (the attributable fraction), we can make cross-level comparisons more precise. Whether these estimates should be expected to be equal is something that must be determined in each specific case. In some instances, we may expect a discepancy between micro and macro estimates of effects. One reason for this is that micro studies generally focus upon the direct effect only of the risk factor, such as the effect of unemployment

on suicide risk. A corresponding macro study would relate the unemployment rate to the suicide rate, and the ensuing effect estimate would capture indirect effects as well, that is, the triggering of suicides among those who fear losing their jobs.

Other Correlates of Suicide

Although the focus of this article is on the relationship between alcohol abuse and suicide, it is of interest to assess the relative importance of alcohol compared with other causes of suicide. Another reason for including additional correlates that have been suggested in the literature is that we thereby reduce the risk of omitted variable bias. This is important particularly in the context of ecological data (Hanushek, Jackson & Kain 1974).

DIVORCE

Divorce is a central indicator of social disintegration considering the fact that the family is the most important primary group in our society. (Family integration was also the factor to which Durkheim devoted most space in his work on suicide.)

A large-scale study including individual-level data for 13 industrialized countries reports an elevated suicide risk among divorced persons in all these countries (Danigelis & Pope 1979). Furthermore, ecological studies (Breault 1986; Gove & Hughes 1980; Stack 1980; Trovato 1986) as well as time-series studies (Stack 1981; Trovato 1987) have found relationships between suicide and divorce (or living alone, in the article by Gove and Hughes).

INCOME

In his review of the research, Stack (1982) reports several macrolevel findings of an inverse relationship between income levels and suicide. Stack's interpretation of this is that "money reduces stress, makes life more worthwhile, and hence reduces the propensity toward suicide" (48).

RELIGIOSITY

According to Durkheim, religious belief protects individuals against suicide by means of its integrative effects in the form of shared beliefs and social ties. While Durkheim's hypothesis of denominational differences with respect to suicide has received little support in modern research (for an overview, see Stack 1982), religiosity per se appears to have an impact. In a study by Breault (1986), who also reviews other studies, religiosity proved to be a powerful predictor of the spatial variation in U.S. suicide rates.

UNEMPLOYMENT

The workplace is a vital locus of social ties and shared goals. To become unemployed means that these integrating forces cease to operate, which should increase the risk of suicide. A large number of aggregate-level studies of suicide have examined the role of unemployment and have generally found it to be significantly related to the suicide rate — for reviews, see Stack (1982) and Platt (1984). Of course, we may ask to what degree the aggregate relationship depends on the suicide response among unemployed persons — the relationship may also depend on reactions among other segments of the population, such as those who fear losing their jobs. However, microlevel evidence from several studies also suggests an increased suicide risk among unemployed people (Iversen et al. 1987; Moser et al. 1987; Norström 1988c; Platt 1984).

URBANIZATION

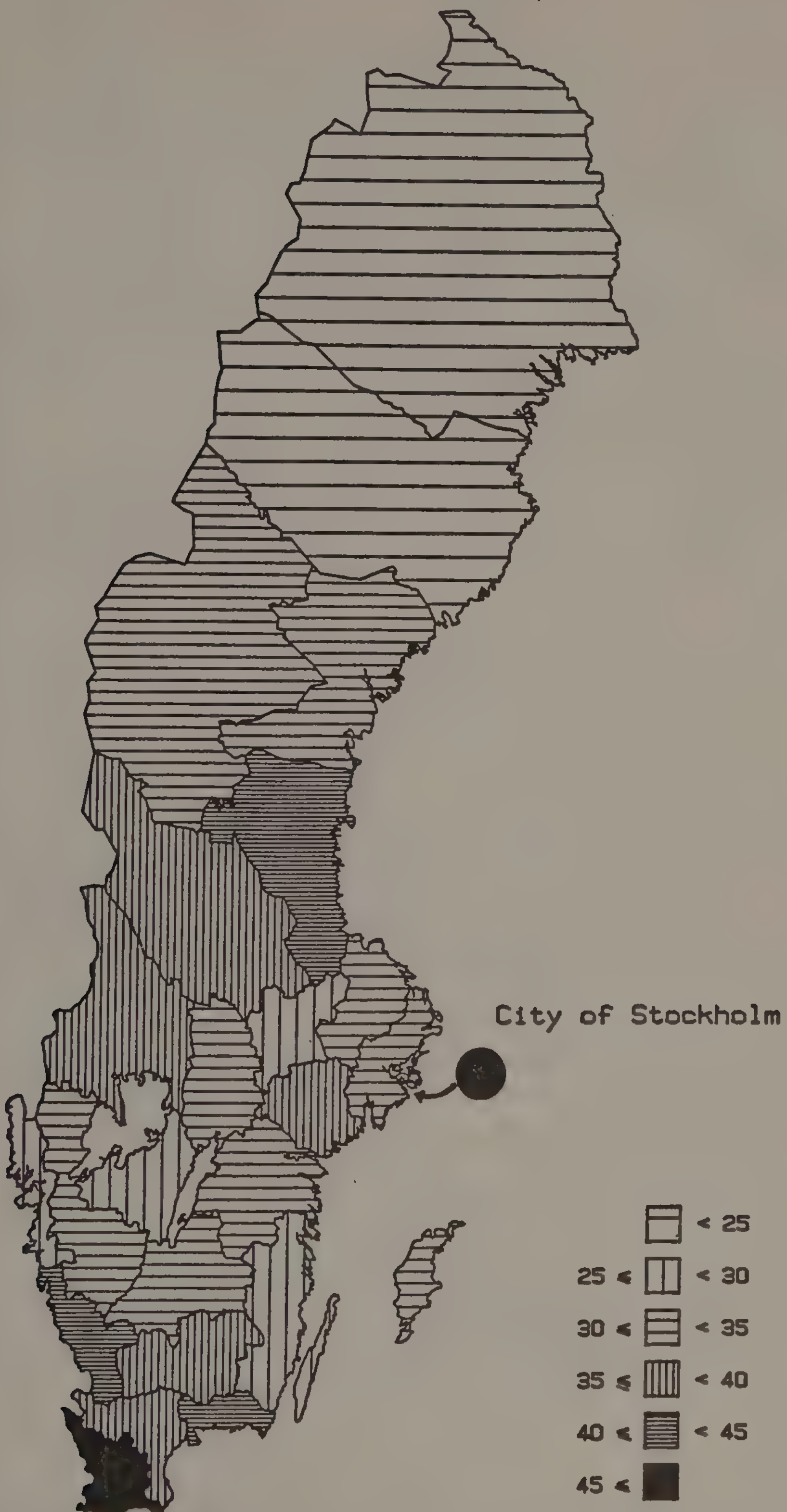
Due to cultural heterogenity and the influx of migrants, urbanized areas can be expected to have a higher degree of disintegration, which in turn should increase the suicide risk. Although the empirical evidence on the link between urbanization and suicide is somewhat inconclusive (see Stack's (1982) review of findings), it seems natural to include urbanization as a control variable in the ecological data.

The Ecological Data

The data points for the dependent variable consist of period averages of the age-standardized male suicide rate (1963-65) in each of Sweden's 24 counties and the city of Stockholm. The reason for averaging over a three-year period is to reduce the stochastic variation that is due to small absolute numbers. As detailed below, the explanatory variables are either averaged in the same manner, or refer to the midpoint of the study period. The delimitation of the study period was determined by the scarcity of county data prior to 1963. Data for more recent time periods are less feasible due to the new code that was added to the cause of death classification in 1969 (ICD AE 149): "uncertainty whether the death was accidental or intentional." A study by Bolander and Ettlinger (1980) suggests that a certain number of actual suicides is concealed in the new code, and that the use of it varies among regions, a fact of great significance in the present context. The ecological pattern of the suicide rate is thereby more affected by "noise" after 1968 which makes it less suitable for analyses.

The spatial distribution of the male suicide rate averaged over the years 1963 through 1965 is depicted in Figure 2. The highest rate is observed for the city of Stockholm (about 40% higher than the national rate), while the two northernmost counties score the lowest (about 50% below the national rate). There is no marked geographical gradient, however. We are still left with a related question: is the suicide rate spatially autocorrelated, that is, is there any tendency for proximate counties to exhibit similar rates? (See, e.g., Cliff & Ord 1973 for a discussion of spatial autocorrelation.) One way of determining this is

FIGURE 2: Male Suicide Rate, Swedish Counties, 1963-1965



to order the data pairwise, each pair consisting of counties with a common border, and then to compute the correlation within these pairs. Excluding the island county of Gotland, there are 48 such pairs. The resulting correlation is .18, which indicates that there is virtually no spatial clustering of this kind in the suicide rate. This may seem somewhat surprising; a cursory glance at Figure 2 gives the impression of some structure. However, closer examination reveals several cases of neighboring counties with greatly differing rates. This result is not only of descriptive value; the presence of spatial autocorrelation would have complicated the regression analysis (Cliff & Ord 1973).

In an analysis of the regional variation in the suicide rate, it is warranted to estimate the stochastic component of this variation first. It appears that the observed variance in suicides is about 16 times greater than should be expected by chance, assuming a Poisson process. (The difference is highly significant: $\chi^2 = 5{,}010.96$, 24 df, $p < .001$.)

We now proceed to the indicators of the explanatory variables.

Per capita consumption of alcohol (liters 100% per capita [15+] averaged over the years 1963-65) is used as an indicator of alcohol abuse. The rationale for this choice is the close association between per capita consumption and heavy drinking, or alcohol abuse, which has been documented in several studies (Skog 1985a, 1985b).

The divorce indicator is the percentage of men (15+) who were divorced (the stock) in 1964.

As pointed out by Breault (1986), the ideal measure of religious integration should include not only affiliation or membership, but also a more active component, i.e., degree of religious involvement and commitment. Such data are as unavailable for Sweden as for the U.S., and we must make the best of what exists. It should be noted that in addition to the State church, there are a number of Free Churches in Sweden (among which the Pentecostal Movement is the largest). Our measure should include both categories since they constitute alternative outlets for religious activity. However, membership in a Free Church reflects active religious affiliation while State church membership is a default option, with each citizen automatically a member unless he or she resigns (an option that few make use of). In light of these considerations, the proxy for religiosity to be used is an additive index comprised by (1) the percentage of the county population who were Free Church members, and (2) religious attendance in the State Church (measured as the annual percent average of the county population attending religious services). The index refers to the midpoint of the period under study.

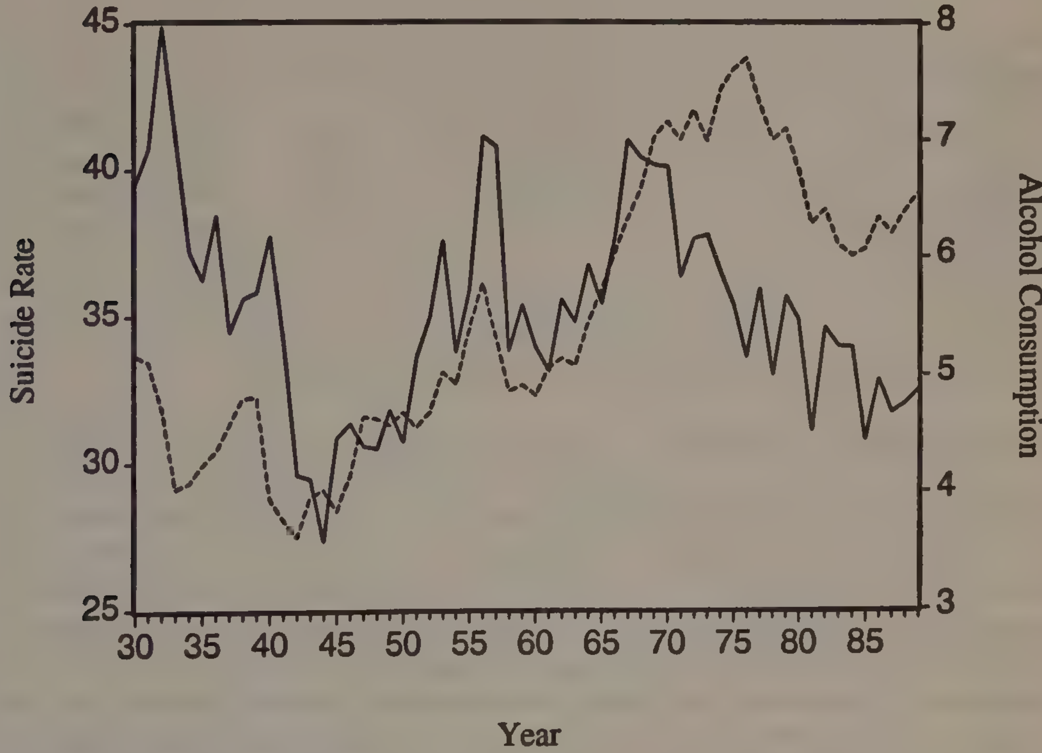
The unemployment measure (the percentage of the labor force who were unemployed) is averaged over the years 1963-65.

The urbanization measure (the percentage of the population living in densely populated areas) refers to the midpoint of the study period.

The Time-Series Data

Swedish time-series data are available for the following of the explanatory variables discussed above: alcohol consumption, divorce, and unemployment. These variables, as well as the suicide rate, are measured in the same way as

FIGURE 3: Male Suicide Rate and Alcohol Consumption (Liters 100%) per Capita, Sweden, 1930-1989[a]



[a] Male suicide rate — solid line
Alcohol consumption (liters 100%) per capita — broken line

their counterparts in the ecological data set. I also consider an alternative divorce indicator, the divorce rate (the incidence of divorces per 100,000 married men). In addition, the time-series data include a measure of real wages (for male industrial workers).

Figure 3 displays the suicide rate and the explanatory variable of primary interest here, alcohol consumption. Although the two series coincide fairly well, graphical evidence of this kind is not very weighty. Particularly when the data are nonstationary, it is crucial to difference the series before performing the analysis, so as to reduce the risk of spurious relationships due to common trends.

Analysis

This section first presents the analyses of the two aggregate data sets. Next, some findings from existent individual-level studies are presented. Finally, we

TABLE 1: Zero-Order Correlations between Male Suicide Rate and Five Explanatory Variables[a]

	X2	X3	X4	X5	X6	$\bar{X}$	Std. Dev.
X1 Ln suicide	.70***	.53**	-.62**	-.54**	.38	3.51	.24
X2 Alcohol consumption		.82***	-.55**	-.23	.56**	3.41	.82
X3 Divorce			-.57**	-.14	.76***	2.15	.95
X4 Religiosity				.14	-.46*	3.90	1.57
X5 Unemployment					-.34	1.61	1.11
X6 Urbanization						71.66	12.10

(N = 25)

[a] Swedish County Data, 1963-1965

* p < .05 ** p < .01 *** p < .001

observe to what degree the spatial and temporal results are consistent with each other, and how they compare with the individual-level data.

It seems likely that the role of our predictors in a suicidal process is of a reinforcing kind, amplifying the impact of each other and that of other suicidogenic factors. This would make a multiplicative model appropriate. Hence a semilogarithmic model of the following general form was used in the analyses of the ecological and the time-series data:

$$Ln\, S = \beta' X + \epsilon \tag{9}$$

where S is the age-standardized male suicide rate (15+), X is a set of explanatory variables, β is a vector of parameters, and ε is the error term.

ECOLOGICAL DATA

The matrix of correlations for the ecological variables (1963-65) is displayed in Table 1. We note that the suicide rate is fairly strongly correlated with most of the predictors; contrary to expectations, however, its correlation with unemployment is negative. The routine inspection of scattergrams revealed that this is due to an outlier (the county of Norrbotten which has a very high unemployment). This observation was therefore excluded in the subsequent analyses. (The correlation dropped in magnitude [to -.33] and became statistically insignificant [$p = .11$]).

It is further noted that the correlation between two of the explanatory variables, divorce and alcohol consumption, is quite strong ($r = .82$). This is at variance with existing microlevel data, which suggest a fairly weak individual-level correlation, in the range from .1 to .2 (Colett & Lindqvist 1971; Norström 1983). A plausible interpretation of this discrepancy would be that the individual-level correlation is compounded in the aggregation process by the geograph-

TABLE 2: Estimates of Semilogarithmic Model for Suicide[a]

	Coeff.	Std. Error	p
Alcohol consumption	.098	.026	.002
Divorce	.028	.019	.158
Religiosity	-.047	.015	.007
Unemployment	-.050	.031	.128
Urbanization	.0005	.002	.790
(N = 24)			

[a] Estimated on Swedish County Data 1963-65. Ridge regression with k = .85.

ical allocation of people. Heavy drinkers and divorced people tend to live in the same areas, but they do not tend to be the same individuals.

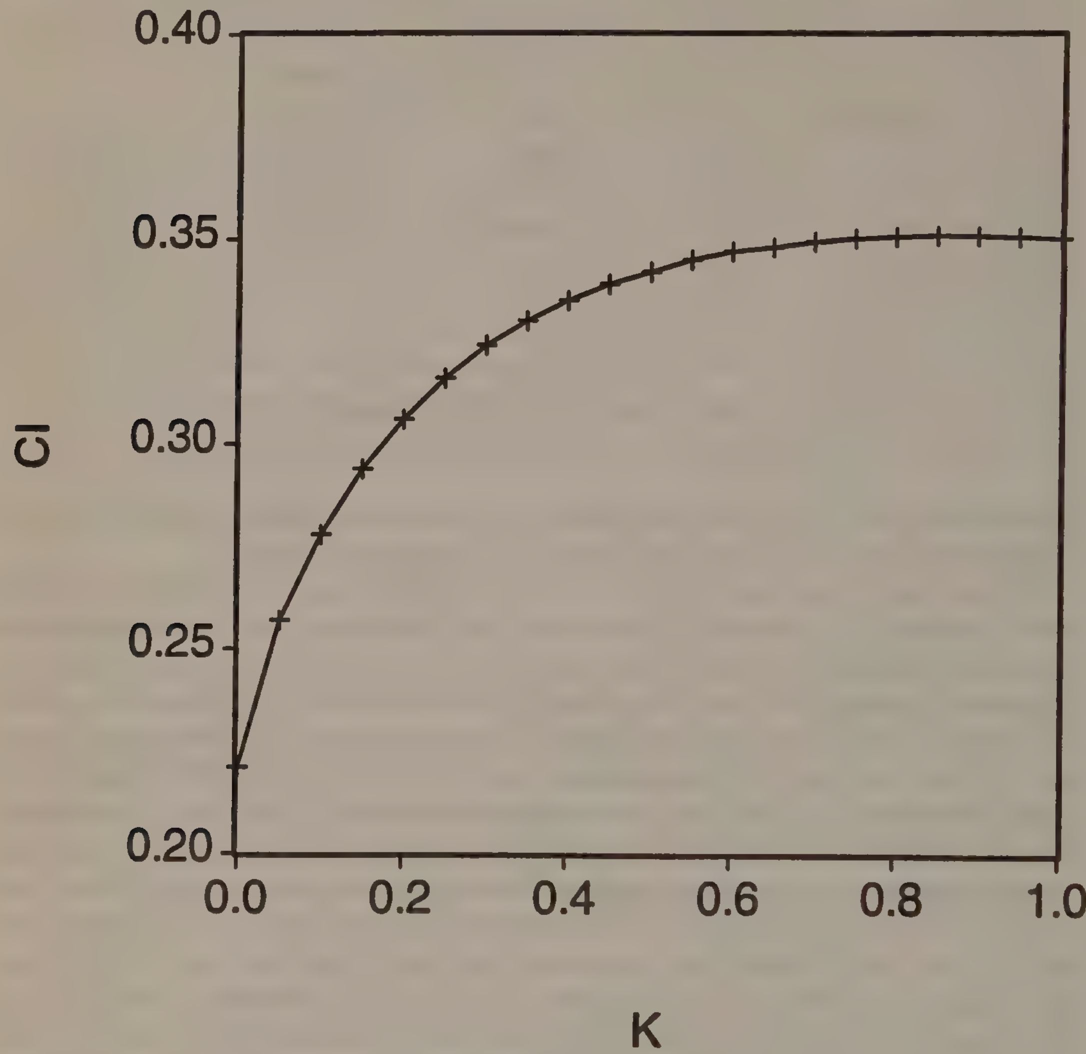
Since the strong correlation between divorce and alcohol consumption would present a problem in ordinary least squares (OLS) regressions, the alternative of ridge regression was applied instead. The idea of this method is to replace the matrix of moments (the $X'X$-matrix) with an $X'X$-matrix in which a constant, k, has been added to the diagonal elements. This is done stepwise, starting with $k = 0$ (which is equivalent to OLS). The loss from the bias thereby introduced is outweighed, up to a certain value of k, by the gain in decreased variance of the parameter estimates, which makes them more efficient and closer to the true values (Draper & Smith 1981; Hoerl & Kennard 1970). A cross-validation index (CI) is used to determine the optimal value of k (Stone 1974). This is constructed as follows: the model is fitted to the data set after deletion of a single data point. From the estimated model parameters we obtain a prediction $\hat{y}_{hi}$ for the deleted observation. The procedure is repeated for each data point, and then we compute:

$$CI = 1 - \frac{\Sigma(y_i - \hat{y}_{hi})^2}{\Sigma(y_i - \bar{y})^2} \tag{10}$$

CI cannot exceed unity but takes a negative value if the explanatory variables are irrelevant.

A semilogarithmic model (equation 9) was fitted to the data, and the ridge constant (k) was fixed to values between 0 and 1. Figure 4 shows the value of CI as a function of k. The problem associated with ridge regression of selecting a feasible value of k (Judge et al. 1980) does not seem very salient in the present analysis. As can be seen, the predictive power associated with OLS ($k = 0$) is much improved, and attains an optimum at $k = .85$. The ensuing parameter estimates are displayed in Table 2. (The coefficients can be interpreted as the percentage change in the suicide rate associated with a one-unit change in the explanatory variable.) It appears that alcohol consumption and religiosity are statistically significant, while the remaining predictors (divorce, unemployment, and urbanization) are much below significance.

FIGURE 4: *CI* as a Function of *k*



TIME-SERIES DATA

The method for analyzing time-series data that has been developed by Box and Jenkins (1976) has several features that render the estimates much more reliable than those that can be expected from ecological analyses. The risk of omitted variable bias is greatly reduced (although not eliminated) by the differencing technique; i.e., the analysis is performed on the yearly changes rather than on the raw data. In this way, linear trends are removed from the data. The rationale for this procedure is that an omitted predictor is more likely to be correlated with an included predictor due to coinciding or diverging trends, rather than due to synchronization in the yearly changes.

Another attractive feature of the Box-Jenkins method is that no assumption of independent error terms is necessary. The noise term (which includes explanatory variables not comprised in the model) is allowed to have a

temporal structure, which is modelled and estimated in terms of autoregressive or moving average parameters.

A semilogarithmic model (equation 9) was fitted to the differenced data for the period 1935-68, which includes the study period of the ecological analysis (1963-65). As a crude test of the presence of structural shifts — which would jeopardize comparisons between the time-series estimates and the ecological estimates — the model was estimated for a longer time period as well; 1930-89. A dummy variable was included to capture the possible effect of the new code (ICD AE 149) that was added to the cause of death classification in 1969 ("uncertainty whether the death was accidental or intentional"). The dummy takes the value of 0 before 1969, and 1 otherwise. The results are shown in Table 3. As can be seen, alcohol consumption and unemployment are statistically significant but divorce and real wages are not. The alternative divorce indicator (the incidence of divorce) also turned out to be statistically insignificant in both time periods, and did not alter the estimates of the remaining predictors (for the longer study period, the t value was -.82 [$p = .42$], and for the shorter study period 1.30 [$p = .21$]). It may also be noted that the differences between the period-specific estimates are negligible; there are thus no indications of structural shifts in parameters. The diagnostic tests of autocorrelated residuals are satisfactory.

INDIVIDUAL-LEVEL DATA

On the basis of microlevel data, it is possible to obtain estimates of attributable fractions with respect to suicide for the following factors: alcohol abuse, divorce, and unemployment.

Alcohol Abuse

There are three retrospective Swedish studies that provide information about the prevalence of alcohol abuse among male suicides (suicide attempters in one case). With material from the 1930s and early 1940s, Dahlgren (1945) found that about 47% of the males who had attempted suicide ($n = 94$) had received the psychiatric diagnosis of alcoholism. In the study by Beskow (1979), comprising 271 male suicides (1970-71), 44% were classified as alcohol abusers on the basis of register documents. The figures reported by Bonnichsen and Geertinger (1969), on 211 male suicides in the 1960s, are less detailed but compatible with Beskow's findings.

It is of course hazardous to draw causal inferences on the basis of the figures cited above; suicide as well as alcohol abuse may well have ensued from common personality factors. However, when reading the case histories of alcoholic suicides (e.g., Murphy 1992), one is struck by the common pattern of how alcohol abuse per se has created problems, or aggravated an otherwise troublesome life situation, with the result that the subject committed suicide. With the absence of alcohol abuse, one can envisage a nonfatal outcome in most cases. Nevertheless it seems quite plausible that the above-cited proportions of alcoholism among suicides are affected by self-selection. An attributable fraction (AF) of about .45 would thus be the upper limit of what should be expected.

TABLE 3: Estimates of Semilogarithmic Model for Suicide[a]

	1935-1968			1930-1989		
	Coeff.	Std. Error	p	Coeff.	Std. Error	p
Alcohol consumption	.100	.040	.019	.109	.028	<.001
Divorce	.062	.289	.832	-.024	.035	.499
Dummy for 1969				-.067	.053	.211
Real wages	.028	.222	.901	-.064	.111	.568
Unemployment	.036	.012	.005	.032	.006	<.001
Noise						
AR(1)	-.389	.177	.036	-.438	.130	.001
AR(2)	-.385	.179	.040	-.327	.129	.014
Diagnostics						
Q	12.82, p > .62			21.92, p > .11		

[a] Estimated on differenced data for Sweden, 1935-1968 and 1930-1989.
Q = Test for residual correlation.

Divorce

According to Swedish suicide data disaggregated by marital status for 1966-1968 (Bolander 1972), the age-standardized relative risk of divorced men with respect to suicide is equal to 2.70 (computed according to formula 4 above). By applying formula 3, we obtain an *AF* equal to .041; about 4% of the male suicides would thus be attributable to divorce. In all probability, this figure is inflated by the divorced being selected with respect to characteristics that are linked to suicide.

Unemployment

A prospective study of unemployed men in Sweden (in the early 1960s) reports an age-standardized relative risk for suicide equal to 3.04, with an associated *AF* of .028 (Norström 1988c). In other words, about 3% of the male suicides would be directly attributable to unemployment. A prospective Danish study (Iversen et al. 1987) reports a relative risk of the same magnitude, 2.92. Since these risk estimates are based on prospective data they should be less vulnerable to selection effects.

The attributable fractions presented above capture only the direct effects on suicide of the risk factors alcohol abuse, divorce, and unemployment. As previously pointed out, the existence of indirect effects appears quite conceivable in some instances, i.e., effects on individuals outside the exposed population. A case in point is unemployment; an elevated suicide risk can be hypothe-

sized not only among those directly affected, but in addition among, e.g., individuals who fear losing their jobs. Contagion effects via social networks is another indirect mechanism that is likely to compound the detrimental impact of job losses (Eyer 1977). Impact estimates based on aggregate data include such indirect effects; a comparison between micro- and macrolevel findings would thus indicate the relative importance of direct and indirect effects.

COMPARISON OF FINDINGS

To what degree are the findings that have been presented above consistent with each other? As regards the alcohol effect, we note that the ecological and the time-series estimates differ very little; a one-liter increase in per capita alcohol consumption is expected to increase the suicide rate by about 10%. The attributable fraction associated with these aggregate findings is 37% (obtained by applying formula 6, and averaging over the ecological and the two time-series estimates). This is somewhat lower than the prevalence estimate of alcoholism among suicides (which was about 45%). The discrepancy suggests that the latter estimate not only reflects the impact of alcohol abuse on the suicide risk but in addition is influenced by self-selection.

The estimated effect of divorce on suicide turned out to be clearly insignificant both in the ecological and in the time-series analysis. It may be recalled that the microlevel evidence here suggested a little less than three-fold increase in suicide risk among divorced men, which implied that about 4% of the suicides would be attributable to divorce. Since this risk estimate is based on cross-sectional data, it is probably inflated by selection effects. The actual risk is thus likely to be even more modest; at any rate it does not appear to be strong enough to show that variations in the number of divorced noticeably affect the aggregate suicide rate.

The aggregate findings are more mixed when we come to unemployment; this variable proved to lack predictive power in the ecological analysis while it was clearly significant in the temporal analysis. Since the latter finding should be the more reliable one, we take it as the point of departure in the cross-level comparison. The attributable fraction that can be derived from the time-series estimates of the unemployment effect on suicide is equal for the two study periods, or .095. This is larger than what was obtained from the prospective micro data (.028). Although the difference should not be overemphasized, it supports the hypothesis discussed earlier of indirect effects of unemployment on the suicide risk.

Conclusions

In this article various data and methods have been triangulated to assess the impact on suicide of a fairly large set of factors, with the emphasis placed on alcohol abuse. To facilitate cross-level comparisons, an epidemiological effect measure, the attributable fraction, has been used. This synthetic approach provides a firmer basis for inferences and has the potential of giving insights about the links between the micro- and macrolevels.

The most unequivocal finding is that of the alcohol effect on suicide risk. Individual-level data, as well as ecological and time-series data yield fairly consistent estimates of this effect, which proves in all instances to be of substantial magnitude; a good third of the male suicides thus appears to be attributable to the alcohol factor. The impact of unemployment also seems to be fairly sizable, although it is substantially smaller than the alcohol effect. It is also less strongly corroborated since it was not sustained by the ecological data. A tentative conclusion here, based on the difference between micro- and macro-level estimates, is that the indirect effects of unemployment on suicide risk are at least as important as the direct ones. As to divorce, individual-level data suggested an effect on the suicide risk. However, this risk was fairly modest when considering the plausible presence of self-selection, and it did not materialize itself at the aggregate level. This outcome is not self-evident; in comparison, it was just noted that the similarly weak inividual-level relationship between unemployment and suicide seems to be compounded at the aggregate level by the operation of indirect effects. Why a similar phenomenon does not occur with respect to divorce is of course hard to ascertain; a possible explanation here is that increases in unemployment are more visible through media exposure and also more tangible due to experiences and discussions at the workplace.

Divorce receives a great deal of attention in suicide research, but our findings suggest that its importance perhaps needs to be reconsidered. Although divorce is a stressful life event, we should recognize that an unhappy marriage is socially and mentally depriving too, implying not only pains but also possible gains from divorce (Renne 1971). In one of the few well-designed studies that has some bearing on this issue (a panel study by Menaghan & Lieberman 1986), the estimated effect of divorce on depression is indeed quite weak (standardized beta = .08). When assessing the existent findings with regard to suicide risk, two qualifications should be borne in mind. First, the microlevel evidence is very likely to be affected by self-selection. In fact, Stafford and Gibbs (1988) found that the individual-level association between marital integration and suicide disappeared after control for other factors. Secondly, the available macrolevel evidence is not very convincing since the analyses typically are performed on trending raw data.

The remaining findings are based on single data sets, but may nevertheless be worth summarizing. Thus, religiosity turned out to be significant, and inversely related to the ecological variation in suicide, while urbanization did not have any predictive power. The same was true for real wages in the temporal analysis of suicide.

It is indeed thought provoking that sociologists studying suicide are so preoccupied with family and work integration variables, but almost totally neglect the alcohol factor. In contrast, psychiatric and epidemiological research has paid close attention to the role of alcohol abuse, supposedly because the diagnosis of alcoholism is so often encountered in the case records. This bias in sociologists' choice of suicide correlates can presumably be attributed to the Durkheimian legacy. Although there is nothing wrong in testing the specific factors that were propounded by Durkheim, deriving new ones from his theory may lead to additional insights. In my view, the alcohol factor is a case in point;

the hypothesis of a relationship between alcohol abuse and suicide fits well into Durkheim's theory of suicide, and it makes a substantial contribution to the explanation of suicidal behavior.

Notes

1. The derivation of formula 8 is as follows. Equation 7 can be written as (suppressing the error term):

$$Y = e^{\alpha} * e^{\beta X}$$

If X is set to 0, then $Y = e^{\alpha}$ Thus, excess in Y due to $X = e^{\alpha} * e^{\beta X} - e^{\alpha} = e^{\alpha}(e^{\beta X} - 1)$. Since AF is defined as the excess due to X, divided by the total, we have:

$$AF = \frac{e^{\alpha}(e^{\beta X} - 1)}{e^{\alpha} * e^{\beta X}}$$

$$= \frac{e^{\beta X} - 1}{e^{\beta X}}$$

$$= 1 - \frac{1}{e^{\beta X}}$$

$$= 1 - e^{-\beta X}$$

2. The occurrence of infrequent events, such as suicide, are normally assumed to be generated by a Poisson process (Feller 1950:153). In a Poisson distribution the mean and variance are identical (Hald 1952:714), and I therefore test by means of the Pearson chi square test whether the observed number of suicides in each county differs from the expected number.

References

Beskow, Jan. 1979. "Suicide and Mental Disorder in Swedish Men." *Acta Psychiatrica Scandinavica*, Supplementum 277.

Bolander, Anne-Marie. 1972. "Nordic Suicide Statistics." Pp. 57-88 in *Suicide and Attempted Suicide*, edited by J. Waldenström, T. Larsson, and N. Ljungstedt. Skandia International Symposia. Nordiska bokhandelns förlag.

Bolander, Anne-Marie, and Ruth Ettlinger. 1980. "Suicide in Sweden." Pp. 39-72 in *Health and Society. V. Different Aspects of Mortality in Sweden*. Committee for Future Oriented Research (Stockholm).

Bonnichsen, Roger, and Preben Geertinger. 1969. "Suicidium i Sverige." (Suicide in Sweden.) *Läkartidningen* 66:2933-36.

Box, George E.P., and Gwilym M. Jenkins. 1976. *Time Series Analysis: Forecasting and Control*. London: Holdens-Day.

Breault, K.D. 1986. "Suicide in America: A Test of Durkheim's Theory of Religious and Family Integration, 1933-1980." *American Journal of Sociology* 92:628-56.

Bruun, Kettil. 1959. *Drinking Behaviour in Small Groups*. The Finnish Foundation for Alcohol Studies.

Bushman, Brad J., and Harris M. Cooper. 1990. "Effect of Alcohol on Human Aggression: An Integrated Research Review." *Psychological Bulletin* 107:341-54.

Cahalan, Don. 1970. *Problem Drinkers*. Jossey-Bass.

Cahalan, Don, and Robin Room. 1974. *Problem Drinking among American Men*. College & University Press.

Cliff, Andrew D., and J. Keith Ord. 1973. *Spatial Autocorrelation*. Pion.

Colett, John, and Gunnar Lindqvist. 1971. "Sambandet mellan Alkoholmissbruk och vissa Sociala Förhållanden." (The Relationship between Alcohol Abuse and Some Social Factors.) Pp. 7:3-7:33 in *Svenska Folkets Alkoholvanor*. Statens Offentliga Utredningar no. 1971:77.

Dahlgren, Karl Gustaf. 1945. *On Suicide and Attempted Suicide*. Lindstedts Universitetsbokhandel.

_____. 1951. "On Death Rates and Causes of Death in Alcohol Addicts." *Acta Psychiatrica Scandinavica* 26:297-311.

Danigelis, Nick, and Whitney Pope. 1979. "Durkheim's Theory of Suicide as Applied to the Family: An Empirical Test." *Social Forces* 57:1080-1106.

DeRicco, Denice A., and Warren K. Garlington. 1977. "The Effect on Modeling and Disclosure of Experimenter's Intent on Drinking Rate of College Students." *Addictive Behaviors* 2:135-39.

Draper, N.R., and H. Smith. 1981. *Applied Regression Analysis*. 2d ed. Wiley.

Durkheim, Emile. [1897] 1966. *Suicide. A Study in Sociology*. Free Press.

Eyer, Joseph. 1977. "Does Unemployment Cause the Death Rate Peak in Each Business Cycle? A Multifactor Model of Death Rate Change." *International Journal of Health Services* 7:625-62.

Feller, William. 1950. *An Introduction to Probability Theory and its Applications*. Vol. 1, 3d ed. John Wiley.

Gove, Walter R., and Michael Hughes. 1980. "Reexamining the Ecological Fallacy: A Study in Which Aggregate Data Are Critical in Investigating the Pathological Effects of Living Alone." *Social Forces* 58:1157-77.

Gustafson, Roland. 1986. *Alcohol and Human Physical Aggression: The Mediating Role of Frustration*. Ph.D. diss. Department of Psychology, University of Uppsala.

Gundelach, Peter. 1992. "Danskernes Intolerance." (The Intolerance of the Danes.) *Dansk Sociologi* 2/92:4-22.

Hald, A. 1952. *Statistical Theory with Engineering Applications*. John Wiley.

Hanushek, Eric A., Jackson, John E., and John F. Kain. 1974. "Model Specification, Use of Aggregate Data, and the Ecological Correlation Fallacy." *Political Methodology* 1:89-107.

Hauge, Ragnar, and Olav Irgens-Jensen. 1986. "The Relationship Between Alcohol Consumption, Alcohol Intoxiction and Negative Consequences of Drinking in Four Scandinavian Countries." *British Journal of Addiction* 81:513-24.

Hoerl, Arthur E., and Robert W. Kennard. 1970. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." *Technometrics* 12:5-67.

Iversen, Lars, Otto Andersen, Per Kragh Andersen, Kristen Christoffersen, and Niels Keiding. 1987. "Unemployment and Mortality in Denmark, 1970-80." *British Medical Journal* 295:879-84.

Judge, George G., William E. Griffiths, R. Carter Hill, and Tsoung-Chao Lee. 1980. *The Theory and Practice of Econometrics*. John Wiley.

Kühlhorn, Eckart, and Mats Ramstedt. 1993. "Tolerans mot Avvikare." (Tolerance toward Deviants.) Department of Sociology, Stockholm University.

Lenke, Leif. 1990. *Alcohol and Criminal Violence — Time Series Analyses in a Comparative Perspective*. Almqvist & Wiksell.

Lieberson, Stanley. 1985. *Making It Count: The Improvement of Social Research and Theory*. University of California Press.

Lilienfeld, Abraham M., and David E. Lilienfeld. 1981. *Foundations of Epidemiology*. Oxford University Press.

Mayfield, Demmie. 1976. "Alcoholism, Alcohol, Intoxication and Assaultative Behavior." *Diseases of Nervous System* 37:288-91.

Menaghan, Elizabeth G., and Morton A. Lieberman. 1986. "Changes in Depression Following Divorce: A Panel Study." *Journal of Marriage and the Family* 48:319-28.

Moser, K.A., P.O. Goldblatt, A.I. Fox, and D.R. Jones. 1987. "Unemployment and Mortality: Comparison of the 1971 and 1981 Longitudinal Study Census Samples." *British Medical Journal* 294:86-90.

Murphy, George E. 1992. *Suicide in Alcoholism*. Oxford University Press.

Murphy, George E., and Eli Robins. 1967. "Social Factors in Suicide." *Journal of the American Medical Association* 199:303-308

Norström, Thor. 1983. "Law Enforcement and Alcohol Consumption Policy as Counter-Measures against Drunken Driving: Possibilities and Limitations." *Accident Analysis and Prevention* 15:513-21.

____. 1988a. "Alcohol and Suicide in Scandinavia." *British Journal of Addiction* 83:553-59.

____. 1988b. "Deriving Relative Risks from Aggregate Data. 1: Theory." *Journal of Epidemiology and Community Health* 42:333-35.

____. 1988c. "Deriving Relative Risks from Aggregate Data. 2: An Application to the Relationship between Unemployment and Suicide." *Journal of Epidemiology and Community Health* 42:336-40.

Nørvig, Johannes, and Børge Nielsen. 1956. "A Follow-Up Study of 221 Alcohol Addicts in Denmark." *Quarterly Journal of Studies on Alcohol* 17:633-42.

Philips, Lorna A. 1976 "An Application of Anomy Theory to the Study of Alcoholism." *Journal of Studies on Alcohol* 37:78-84.

Platt, Stephen. 1984. "Unemployment and Suicidal Behavior: A Review of the Literature." *Social Science & Medicine* 19:93-115.

Renne, Karen S. 1971. "Health and Marital Experience in an Urban Population." *Journal of Marriage and the Family* 33:338-50.

Ritson, Bruce. 1977. "Alcoholism and Suicide." Pp. 271-77 in *Alcoholism. New Knowledge and New Responses*, edited by Edvards, G., and M. Grant. Croom Helm.

Robins, Eli, George E. Murphy, Robert H. Wilkinson, Seymour Gassner, and Jack Kayes. 1959. "Some Clinical Considerations in the Prevention of Suicide Based on a Study of 134 Successful Suicides." *American Journal of Public Health* 49:888-99.

Roslund, Barbro, and Carl A. Larson. 1979. "Crimes of Violence and Alcohol Abuse in Sweden." *International Journal of Addictions* 14:1103-15.

Rossow, Ingeborg. 1993. "Suicide, Alcohol, and Divorce; Aspects of Gender and Family Integration." *Addiction* 88: 1659-65.

Singer, Estelle, Howard T. Blane, and Richard Kasschau. 1964. "Alcoholism and Social Isolation." *Journal of Abnormal and Social Psychology* 69:681-85.

Skog, Ole-Jørgen. 1985a. "The Collectivity of Drinking Cultures: A Theory of the Distribution of Alcohol Consumption." *British Journal of Addiction* 8:83-99.

____. 1985b. *The Distribution of Alcohol Consumption. Part 3: Evidence of a Collective Drinking Culture*. SIFA-report no. 8/85 (Oslo, National Institute for Alcohol Research).

____. 1986. "The Long Waves of Alcohol Consumption: A Social Network Perspective on Cultural Change." *Social Networks* 8:1-32.

____. 1991. "Alcohol and Suicide — Durkheim revisited." *Acta Sociologica* 34:193-206.

____. 1993. "Alcohol and Suicide in Denmark 1911-24 — Experiences from a 'Natural Experiment'." *Addiction* 88:1189-93.

Skog, Ole-Jørgen, and Elisabet Bjørk. 1978. "Alkohol og Voldskriminalitet: En Analyse av Utviklingen i Norge 1931-1982." (Alcohol and Criminal Violence: An Analysis of the Development in Norway 1931-1982.) *Nordisk Tidskrift for Kriminalvidenskab* 75:1-23.

Skog, Ole-Jørgen, and Zusanna Elekes. 1993. "Alcohol and the 1950-90 Hungarian Suicide Trend — Is There a Causal Connection?" *Acta Sociologica* 36:33-46.

Stack, Steven. 1980. "The Effect of Marital Dissolution on Suicide." *Journal of Marriage and the Family* 42:83-92.

____. 1981. "Divorce and Suicide: A Time Series Analysis, 1933-1970." *Journal of Family Issues* 2:77-90.

____. 1982. "Suicide: A Decade Review of the Sociological Literature." *Deviant Behavior* 4:41-66.

Stafford, Mark C., and Jack P. Gibbs. 1988. "Change in the Relation between Marital Integration and Suicide Rates." *Social Forces* 66:1060-79.

Stone, Mervyn. 1974. "Cross-Validatory Choice and Assessment of Statistical Predictions (with Discussion)." *Journal of the Royal Statistical Society* (Series B) 36:111-47.

Straus, Robert, and Raymond G. McCarthy. 1951. "Nonaddictive Pathological Drinking Patterns of Homeless Men." *Quarterly Journal of Studies on Alcohol* 12:601-11.

Sundby, Per. 1967. *Alcoholism and Mortality*. Almqvist & Wiksell.

Trovato, Frank. 1986. "The Relationship between Marital Dissolution and Suicide: The Canadian Case." *Journal of Marriage and the Family* 48:341-48.

____. 1987. "A Longitudinal Analysis of Divorce and Suicide in Canada." *Journal of Marriage and the Family* 49:193-203.

Wasserman, Ira M. 1989. "The Effects of War and Alcohol Consumption Patterns on Suicide: United States, 1910-1933." *Social Forces* 68:513-30.

White Male Suicide in the United States: A Multivariate Individual-Level Analysis*

AUGUSTINE J. KPOSOWA, *University of California, Riverside*
K.D. BREAULT, *Austin Peay State University*
GOPAL K. SINGH, *National Center for Health Statistics*

Abstract

Using data from the 1979-85 National Longitudinal Mortality Study and multivariate hazards regression analysis, the study investigates risk factors associated with suicide mortality among white males in the U.S. (ICD-9 Codes E950-E959). Results were mixed with regard to the social integration-suicide hypothesis. Divorced or separated men and those who live alone (socially isolated) have significantly higher risks of suicide mortality. However, single and widowed men do not have a significantly greater suicide risk after controlling for key social factors such as socioeconomic status (SES). The results also show that men who live in urban areas and those who are native born are at a higher risk for suicide.

Are marital status and social isolation related to the risk of suicide? In sociology, the question has almost exclusively been investigated at the aggregate level, with the finding that marriage and other sources of social integration reduce the probability of suicide. In psychology and research psychiatry, similar findings have been obtained at the individual level but with insufficient controls for key social variables (e.g., income, education, occupational status, etc.). There are no methodologically rigorous studies at the individual level in the sociology literature, and no studies in the psychological literature with proper social controls. Using data from the 1979-85 Longitudinal Mortality Study, we employ multivariate hazards regression to investigate risk of suicide at the individual level with controls for social integration, place of residence, age, nativity, education, occupation, and family income.

* *The article has benefited from the help of the anonymous reviewers.* *Direct correspondence to Augustine Kposowa, Department of Sociology, University of California, Riverside, CA 92521-0419.*

Theoretical Issues

While the theoretical underpinnings of this research are familiar to most readers, a brief statement may be in order. A long tradition in social thought, perhaps the most notable representative of which is Durkheim, has posited a relationship between suicide and social integration. Marriage and other social bonds are said to be protective against suicide, while the absence of social bonds is seen as causative of suicide. For example, in talking about egoistic suicide, Durkheim argued that married persons have lower suicide rates than singles and widowers. A truly vast amount of research has been done on the general social integration-suicide thesis (among the names most often cited in sociology are Breault, e.g., 1986; Danigelis & Pope 1979; Lester, e.g., 1983; Pescosolido & Georgiana 1989; Stack, e.g., 1982, 1985; Wasserman, e.g., 1984). Sociologists have generally supported the theory, but with few exceptions the research has been based on aggregates such as countries, states, standard metropolitan statistical areas (SMSAs), counties, cities, etc. On the other hand, psychological writers have generally produced the same results at the individual level, but they have done so without controls for key social variables such as income, education, and occupational status (e.g., Goldney 1981; Humphrey 1977; Robins et al. 1977; Shafii et al. 1985; Teicher & Jacobs 1966). The question we ask is simply: can the theory be sustained at the individual level when sufficient social controls are in place?

Two concerns led directly to this research. First, as mentioned above, the individual-level research in psychology and psychiatry fails to address key sociological variables, and second, there is evidence to suggest that the aggregate level findings in sociology may be artifactual (Breault 1995). Breault (1995) has shown that ecological correlates of suicide generally become insignificant once geographical region is controlled. Specifically, there appears to be no relationship between suicide and social integration in either the eastern or western region of the U.S., nor at the level of the state. Prior research findings were contingent on the east-west gradient in suicide and social integration (i.e., high suicide and low social integration in the west and vice versa). Such work cannot be regarded as conclusive because it is conceivable that national results may be correct even though regional and state-level results show the opposite (consistent findings from one level to another and, failing that, lower level results are the preference), but it raises the possibility that at the aggregate level suicide and social integration are fortuitously related. Because the issue cannot be adequately resolved at the group level, we must turn to the individual.

Note that, in contrast to the present research, Durkheim's work is at the aggregate level. There are those (beginning with Parsons) who argue that the study of individuals is irrelevant to Durkheim's theory. We see two problems here. First, as suggested above, the literature is at a point where definitive results cannot be obtained at the aggregate level (and this applies to purely aggregate-level hypotheses, that is, no individual-level predictions). Therefore, if we accept the traditional view, Durkheim's theory is now not falsifiable. Second, in this context Durkheim's views have not been fully understood. His primary interest is in a quantitative sociology in which theoretical ideas are

empirically adjudicated. His hopes were not for a sociology shielded from empirical scrutiny by the notion of aggregation, but the very opposite — a sociology that could compete empirically and triumph over other social sciences because of the superior truth value of its theory. Today, Durkheim would not sacrifice the scientific or theoretical status of his work in favor of a rigid allegiance to the aggregation concept. To do so would jeopardize his very notion of, and ambition for, sociology. For us to do so today only helps to isolate sociology methodologically from the other social sciences.

Materials and Methods

DATA

The data used for the study were drawn from the National Longitudinal Mortality Study (NLMS). The NLMS, a long-term prospective study of mortality in the U.S., was conducted by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the U.S. Bureau of the Census and the National Center for Health Statistics (Rogot et al. 1992). The main objective of the project is to examine socioeconomic, demographic and occupational disparities in mortality in the U.S.

The public-use data file used in our analysis consisted of a cohort of five national samples from the Current Population Surveys between 1979 and 1981. Survival (mortality) experiences were studied until 1984-86 (NHLBI 1992). Data on mortality were collected by using the death certificate information available from the National Death Index (NDI), which were appended to the socioeconomic and demographic characteristics of the 1979-81 population cohort.

The NDI was originally designed to provide a simplified method of identifying the mortality experiences of participants in prospective studies. It is a centralized and computerized index of death records beginning with deaths in 1979 (Rogot et al. 1992). The NDI is maintained by the National Center for Health Statistics using information provided by state vital statistics offices.

The original study population for the NLMS comprised eight census samples numbering about one million people of all ages and races. The samples were mainly drawn from the Current Population Survey and are designated as "cohorts" for mortality follow-up. The individuals were known to be alive on the survey date and therefore eligible for follow-up with regard to survivorship from that date. Of the eight cohorts, seven were derived from the Current Population Surveys, and one was drawn from the 1980 census (Kposowa & Singh 1994; Rogot et al. 1992).

For purposes of this analysis, we considered only white males aged 15 years and older, because the number of suicide deaths among other race groups and women was too small to permit meaningful analysis. Note that suicide deaths among white males constitute nearly 75% of all suicide deaths that occur in the U.s. (CDC 1985). Moreover, suicide was the seventh leading cause of death for white males in 1989, but it was not among the leading causes of death for other race-sex subgroups of the population (NCHS 1992). The sample initially consisted of 203,437 white males which 216 had committed suicide by the end of the five-year follow-up.

VARIABLES

The dependent variable in the study was risk of suicide mortality. Deaths resulting from suicide were defined according to the Ninth Revision of the International Classification of Diseases (ICD-9), with the underlying cause codes E950-E959.

In estimating suicide risk, all those surviving beyond the five year follow-up period constituted right censored observations. Those who died of causes other than suicide during the five year follow-up period were also treated as right-censored.

The risk of suicide was modeled as a function of social integration indicators and other sociodemographic covariates. Specifically, the covariates employed included place of residence, age, nativity, education, occupation, family income, and social integration. Social integration was the main covariate in the analysis, and following Durkheim we used marital status and social isolation as proxies. Marital status was measured by a set of dummy variables, one each for single, widowed, separated, or divorced status. Those married at the beginning of the study constituted the reference category. Social isolation, as indexed by household size, was measured as a dichotomous dummy variable, with code 1 for those living alone and 0 for those who lived with at least one other person in the same household. We dummy coded our variables because our interest is on interpreting relative risks. The latter are much easier to interpret, more meaningful, and more intuitive when one compares subcategories of a variable to a reference group (Kposowa 1995; Kposowa et al. 1994:280).

The other covariates were measured as categorical variables, with percentage distributions shown in the first column in Table 1. Place of residence was measured by two dummy variables, one for SMSA, central city residence, and the other for SMSA, outside central city residence. Those who resided in a nonmetropolitan area represented the reference category. Age was captured by defining it in terms of a series of dummy vectors, one each for groups 25-34, 35-44, 45-54, 55-64, and 65 years and older. Individuals aged 15-24 years were treated as the reference category. Education was measured by two dummy variables, one for high school education and the other for 13 or more years. Individuals with less than a high school education constituted the reference group. Annual family income (1980 adjusted dollars) was indexed by three dummy variables, one each for $10,000-19,999, $20,000-24,999, and $25,000 and over. Those with family income of less than $10,000 were the omitted group. Occupation was represented by two dummy variables; one for those with nonprofessional occupations, and the other for the unemployed and those outside the labor force. Managerial and professional specialty occupations, coded 0, constituted the reference category. Admittedly crude, the professional-nonprofessional categories of occupational status are the only ones available with these data. Nativity was measured as a dichotomous dummy variable, coded 1 for native-born (those born in the U.S.). The reference category, coded 0, included the foreign-born.

THE STATISTICAL MODEL

The Cox proportional hazards model was used to estimate the effects of social integration variables on the risk of suicide mortality (Cox 1972; Gill 1992; Kposowa et al. 1994). The Cox model may be specified as,

$$h(t) = h_0(t)\exp(\Sigma_k B_k X_k) \tag{1}$$

where $h(t)$ is the hazard or risk of suicide mortality at time t, β_k are a set of unknown parameters to be estimated and X_k are K covariates. $h_0(t)$ is a baseline hazard function and is defined when all the covariates in the model assume the value zero. This model is called "proportional hazards" because for any two individuals at time t with hazards $h_i(t)$ and $h_j(t)$, the ratio of their hazard functions

$$\frac{h_i(t)}{h_j(t)} = \frac{h_0(t)\exp(\beta X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik})}{h_0(t)\exp(\beta_1 X_{j1} + \beta_2 X_{j2} + \ldots + \beta_k X_{jk})} \tag{2}$$

$$= \exp[\beta_1(X_{i1} - X_{j1}) + \beta_2(X_{i2} - X_{j2}) + \ldots + \beta_k(X_{ik} - X_{jk})]$$

is a constant, which may depend on the covariates, but not on survival time t (Singh & Kposowa 1994). The basic assumption of the model is that the covariates considered capture population heterogeneity, and that the hazard ratios remain unchanged over survival time.

Let T be a nonnegative random variable representing the elasped time from start of follow-up until the occurrence of suicide death. Note that the survival time, measured in years, as generated by each individual followed and each suicide, comprises the response variable in the present study, and it is inversely related to the hazard rate $h(t)$. In the case of the Cox model, the survival distribution function or survivor function of T, defined as the probability of surviving from follow-up to at least time t, is given by

$$=S(t) = Pr(T \geq t) = [S_o(t)]\exp(\sigma_k \beta_k X_k) \tag{3}$$

where $S_o(t)$ is the baseline survivor function representing the reference categories of the covariates. The proportionality of hazards was tested by inspecting the plots of $ln(-ln[S(t)])$ against survival time t for the various covariate categories, including those for marital status and social isolation. The plots were found to be approximately parallel, and hence the proportionality assumption was taken to be satisfied by the data.

The parameters in the hazards model as specified in equation 1 were estimated by partial likelihood using the PHREG procedure available within SAS (1991). The mathematical details of partial likelihood are discussed elsewhere (e.g., Allison 1984; Kalbfleisch & Prentice 1980; Namboodiri & Suchindran 1987).

Estimated relative risks were obtained by exponentiating the hazard coefficients ($\beta_k s$) in equation 1. Thus, the relative risk (RR) = $\exp(\beta_k)$. An RR greater than one indicates that the risk of suicide death is higher for the covariate category compared to the reference group. An RR of less than one has the opposite interpretation.

TABLE 1: Bivariate Hazards Regression Estimates of the Effects[a] of Sociodemographic and Social Integration Factors on Suicide Mortality for White Males in the U.S., 1979-1985

Covariate	Percent[b]	β	Std. Error (β)	RR	Model χ^2	df
Marital status					35.25**	8
Married	65.4			1.00		
Single	26.6	0.477*	.223	1.61		
Divorced/separated	5.9	0.974**	.218	2.65		
Widowed	2.2	0.677	.362	1.97		
Social isolation					32.75**	6
Lives with others	91.6			1.00		
Lives alone	8.4	0.858**	.183	2.36		
Place of residence					19.52**	7
Nonmetropolitan	40.9			1.00		
SMSA, central city	22.3	0.393*	.173	1.48		
SMSA, noncentral city	36.9	0.185	.160	1.20		
Age					14.44**	5
15-24	23.7			1.00		
25-34	22.1	-0.445*	.223	0.64		
35-44	15.4	0.011	.221	1.01		
45-54	13.4	-0.033	.234	0.97		
55-64	12.7	0.332	.215	1.39		
65+	12.8	0.344	.221	1.41		
Nativity					19.06**	6
Foreign-born	94.9			1.00		
Native-born	5.1	0.846	.453	2.33		
Education					16.73**	7
< 12 years	33.6			1.00		
12 years	33.1	-0.249	.170	0.78		
13+ years	33.4	-0.175	.160	0.84		
Occupation					16.18**	7
Professional	22.7			1.00		
Nonprofessional	55.3	0.212	.186	1.24		
Unemployment/NLF	22.0	0.281	.236	1.32		
Family income					19.94**	8
< $10,000	25.0			1.00		
$10,000-19,999	29.9	-0.221	.176	0.80		
$20,000-24,999	14.6	-0.336	.229	0.71		
$25,000+	30.5	-0.428*	.188	0.65		
(N [all models] = 203,437)						

[a] Adjusted for the effects of age only.
[b] Percent of population in each category.

* $p < .05$ ** $p < .01$

The statistical significance of the effect of each model covariate was assessed by the *t* statistic, obtained as a ratio of the parameter estimate to its corresponding asymptotic standard error.

Results

BIVARIATE ANALYSIS

In Table 1, we present the gross effects of each of the covariates on the risk of suicide mortality, adjusting only for the effect of age. We estimated a separate hazards regression model for each covariate to determine the crude but age-adjusted effect on suicide mortality. The estimated hazard coefficient (β) denotes the expected difference in the log of the suicide hazard between the covariate category (included in the model) and the reference group, controlling for age. Since β is somewhat awkward to interpret as the incremental net impact of the covariate on the log of the suicide hazard, we rely on the measure, $RR = \exp(\beta)$, to provide an intuitively more meaningful interpretation of the effect of the covariate on suicide mortality (Kposowa 1995).

For each covariate, the category showing an *RR* of 1.0 represents the reference group for comparison. The model chi-square (the likelihood-ratio statistic), which is a "global" test statistic for the whole model, was statistically significant at the .01 level for all models in Table 1, indicating that each of the covariates in the model significantly improved the fit of the model.

Most of the results in Table 1 are consistent with expectations concerning the relationships of the covariates and suicide mortality. As may be seen from the table, the risk (hazard) of suicide was 2.65 times as much for the divorced and separated as for the married. Moreover, being single was associated with significantly higher risk of suicide death. Compared to married individuals, the risk of suicide mortality was 61% greater among the single. Widowed status was not significantly related to suicide risk.

With regard to social isolation, persons living alone were more than twice as likely (2.36) to commit suicide than those living with others.

Consistent with results from some ecological studies, living in an SMSA (within the central city), is associated with a significantly higher risk of suicide than living in a nonurban environment. Specifically, the risk of suicide for suburban and central city residents exceeded that for nonurban residents by 20% and 48% respectively.

Compared to those aged 15-24 years, those in the 25-34 age group had a significantly lower risk of suicide. However, the relative risk of suicide was approximately 2.2 times as high for those 55 years and older as for those aged 25-34 years.

Although higher SES, as represented by education, occupation, and family income, was generally associated with lower risk of suicide, only those with family incomes less than $10,000 and those with more than $25,000 showed a significant difference in relative suicide risk.

The estimated relative risk for nativity shows a very strong effect of nativity status on suicide mortality. Those born in the U.S. experienced a risk of suicide that was 2.33 times as large as that for those born outside the U.S.

TABLE 2: Multivariate Hazards Regression Estimates of the Net Effects of Marital Status and Social Isolation on Suicide Mortality for White Males in the U.S., 1979-1985

	Model 1			Model 2		
Covariate	β	Std. Error (β)	RR[a]	β	Std. Error (β)	RR[b]
Marital status						
Married			1.00			
Single	0.389	.222	1.48			
Divorced/separated	0.858**	.221	2.36			
Widowed	0.613	.363	1.85			
Social isolation						
Lives with others						1.00
Lives alone				0.769**	.190	2.16
Place of residence						
Nonmetropolitan			1.00			1.00
SMSA, central city	0.416*	.176	1.52	0.428*	.176	1.53
SMSA, noncentral city	0.263	.164	1.30	0.280	.164	1.32
Age						
15-24			1.00			1.00
25-34	-0.252	.267	0.78	-0.426	.238	0.65
35-44	0.297	.279	1.35	0.109	.229	1.12
45-54	0.259	.291	1.30	0.047	.239	1.05
55-64	0.584*	.275	1.79	0.337	.217	1.40
65+	0.469	.306	1.60	0.185	.242	1.20
Nativity						
Foreign-born			1.00			1.00
Native-born	0.996*	.456	2.71	1.012*	.456	2.75
Education						
< 12 years			1.00			1.00
12 years	-0.206	.174	1.39	-0.241	.174	0.79
13+ years	-0.078	.191	0.93	-0.136	.192	0.87
Occupation						
Professional			1.00			1.00
Nonprofessional	0.170	.204	1.19	0.206	.204	1.23
Unemployment/NLF	0.115	.257	1.12	0.185	.257	1.20
Family income						
< $10,000			1.00			1.00
$10,000-19,999	-0.175	.180	0.84	-0.158	.181	0.85
$20,000-24,999	-0.269	.234	0.76	-0.238	.237	0.79
$25,000+	-0.359	.200	0.70	-0.302	.205	0.74
Model χ^2		51.28**			49.76**	
Degrees of freedom		18			16	
(N = 203,437)						

[a] Adjusted for all covariates except social isolation.
[b] Adjusted for the effects of all covariates except marital status.

* p < .05 ** p < .01

MULTIVARIATE ANALYSIS

Table 2 presents the net effects of marital status and social isolation on the risk of suicide mortality, controlling for the effects of the other covariates. Because they are highly correlated, inclusion of both marital status and social isolation in the same model reduces the importance of the two covariates. Those who are single, divorced, separated, or widowed are quite likely to live by themselves, whereas married individuals by definition are most likely to live with at least one other person. Therefore, in order to assess the net effect of marital status on suicide mortality, we estimated a hazards model that included all the covariates except social isolation (see model 1, Table 2). Similarly, to determine the net effect of social isolation on suicide mortality, we estimated a hazards model with all the covariates except marital status (model 2, Table 2).

As can be seen from model 1 of Table 2, when the potentially confounding effects of all other covariates were taken into account, the excess suicide mortality associated with being single, divorced, separated or widowed was significantly reduced. Compared to their married counterparts, those who were divorced or separated had a 136% higher risk of suicide. However, single and widowed status was not significantly related to suicide risk. For social isolation, while the net effect (2.16 — reported in model 2 of Table 2) was somewhat smaller than the gross effect (2.36) shown in Table 1, living alone still contributed significantly to higher risk of suicide in comparison to living with others.

Because young adults are less likely to be married, we eliminated those in the sample younger than 18. Significant differences did not obtain.

Finally, a full model incorporating all the risk factors (simultaneously controlling for marital status and social isolation), indicated that living in the central city of an SMSA was still associated with significantly higher risk of suicide mortality than living in a nonmetropolitan area. Specifically, compared to nonurban residents, those living in the central city of an SMSA were 50% more likely to commit suicide. This result can be taken as support for Durkheim's theory as many authors, (for example, Breault & Kposowa 1987) see the central city as less socially integrated than nonurban areas.

An additional finding that merits attention is that the native-born had significantly higher risk of suicide than the foreign-born. Even when all other covariates were controlled, nativity status remained a strong predictor of suicide mortality. Those born in the U.S. had almost three times as high a risk of suicide as the foreign-born. Future research may help to unravel the empirical association by differentiating the foreign-born by country of origin. It would be of interest to see if persons emigrating from countries with high rates of suicide exhibit especially low risk of suicide.

Summary and Conclusion

In sociology, previous research on the hypothesized relationship between social integration and suicide has been based largely on aggregates such as countries, states, SMSAs, counties and cities. Few sociologists have analyzed data at the individual level and fewer still have used rigorous methods. Moreover, there is

evidence that the prior aggregate-level research fortuitously linked social integration and suicide, that the relationship disappears once geographic region is controlled. In psychology, while there are several methodological concerns, the major problem is that researchers have failed to include important sociological factors such as SES. In the present study, using a national longitudinal sample of U.S. white males, and controlling for key sociological factors, we find mixed support for the social integration hypothesis: suicide is significantly related to married status in the expected negative direction and positively related to social isolation, but, in a direct contradiction of Durkheim, single and widowed status are unrelated to suicide risk.

Obviously, much more work needs to be done. In looking toward the future, we would invite other suicide researchers to join us on the frontier of individual-level studies. After all, as we have shown, our lives may depend on it.

References

Allison, Paul D. 1984. *Event History Analysis*. Sage

Breault, K.D. 1995. "Was Durkheim Right?: A Critical Survey of the Empirical Literature on *Le Suicide*." Pp. 35-62 in *Celebrating the Centennial of Durkheim's* Le Suicide, edited by David Lester.

Breault, K.D. 1986. "Suicide in America: A Test of Durkheim's Theory of Religious and Family Integration, 1933-1980." *American Journal of Sociology* 92:628-56.

Breault, K.D. and Augustine J. Kposowa. 1987. "Explaining Divorce in the United States: A Study of 3,111 Counties, 1980." *Journal of Marriage and the Family* 49:549-58

Centers for Disease Control. 1985. *Suicide Surveillance Report, United States, 1970-1980*. U.S. Department of Health and Human Services. Public Health Service.

Cox, D.R. 1972. "Regression Models and Life-Tables (with discussion)." *Journal of the Royal Statistical Society* (B)34:187-220.

Danigelis, Nick and Whitney Pope. 1979. "Durkheim's Theory of Suicide as Applied to the Family: An Empirical Test." *Social Forces* 57:1081-1106.

Gill, Richard D. 1992. "Multistate Life-Tables and Regression Models." *Mathematical Population Studies* 3:259-76.

Goldney, R.D. 1981. "Parental Loss and Reported Childhood Stress in Young Women Who Commit Suicide." *Acta Psychiatrica Scandanavica* 64:34-59.

Humphrey, J. 1977. "Social Loss." *Diseases of the Nervous System* 38:157-60.

Kalbfleisch, John D., and Ross L. Prentice. 1980. *The Statistical Analysis of Failure Time Data*. John Wiley.

Kposowa, Augustine J. 1995. "Immigrants and Economic Dependence: Approaches to Presenting Logistic Regression Results." *Applied Behavioral Science Review* 3.

Kposowa, Augustine J., and Gopal K. Singh. 1994. "The Effect of Marriage on Male and Female Homicides in the United States." *Sociological Focus* 27:343-62.

Kposowa, Augustine J., Gopal K. Singh, and K.D. Breault. 1994. "The Effects of Marital Status and Social Isolation on Adult Male Homicides in the United States." *Journal of Quantitative Criminology* 10:277-89.

Namboodiri, Krishnan, and C.M. Suchindran. 1987. *Life Table Techniques and their Applications.* Academic Press.

National Center for Health Statistics. 1992. "Advance Report of Final Mortality Statistics, 1989." *Monthly Vital Statistics Report* 40(8):19-25, Supplement 2.

National Heart, Lung, and Blood Institute. 1992. *National Longitudinal Mortality Study, Public Use File Documentation.* National Institutes of Health.

Pescosolido, Bernice A., and Sharon Georgianna. 1989. "Durkheim, Suicide, and Religion: Toward a Network Theory of Suicide." *American Journal of Sociology* 54:33-48.

Robins, L.N., P.A. West, and G.E. Murphy. 1977. "The High Rate of Suicide in Older White Men: A Study Testing Ten Hypotheses." *Social Psychiatry* 12:1-20.

Rogot E., P.D. Sorlie, N.J. Johnson, and C. Schmitt. 1992. *A Mortality Study of 1.3 Million Persons by Demographic, Social, and Economic Factors: 1979-1985 Follow-Up.* National Institutes of Health. NIH Publication No. 92-3297.

SAS Institute, Inc. 1991. *SAS/STAT Software: The PHREG Procedure.* Cary, North Carolina: SAS Institute.

Shafii, M., S. Carrigan, J.R. Whittinghall, and A. Derrick. 1985. "Psychological Autopsy of Completed Suicide in Children and Adolescents." *American Journal of Psychiatry* 142:1061-64.

Singh, Gopal K., and Augustine J. Kposowa. 1994. "A Comparative Analysis of Infant Mortality in Major Ohio Cities: Significance of Socio-Biological Factors." *Applied Behavioral Science Review* 2:77-94

Stack, Steven. 1982. "Suicide: A Decade Review of the Sociological Literature." *Deviant Behavior* 4:41-66.

____. 1985. "The Effect of Domestic/Religious Individualism on Suicide, 1954-78." *Journal of Marriage and the Family* 47:431-47.

Teicher, J.D., and J. Jacobs. 1966. "Adolescents Who Attempt Suicide: Preliminary Findings." *American Journal of Psychiatry* 122:1248-57.

Wasserman, Ira. 1984. "A Longitudinal Analysis of the Linkage between Suicide, Unemployment, and Marital Dissolution." *Journal of Marriage and the Family* 46:853-59.

Country Music, Suicide, and Spuriousness*

JEFFREY B. SNIPES, *Florida State University*
EDWARD R. MAGUIRE, *State University of New York at Albany*

Abstract

Stack and Gundlach (1992) tested and supported the hypothesis that the greater the air time devoted to country music in metropolitan areas, the greater the white suicide rate will be. We could not replicate this effect (Maguire & Snipes 1994), and Stack and Gundlach (1994) responded with a criticism of our methods, more specifically demonstrating measurement error in our construction of suicide rates. Here we show that this criticism sidesteps our most relevant critique of their study. Yet, we heed their advice, reattempt the replication using their own white suicide data, and still fail to produce a significant effect of country music on white suicide. Although the two are related bivariately, controlling for divorce, poverty, southern region, and gun availability results in a near-zero multivariate effect.

Following the recent wave of research on the relationship between culture and destructive behavior, Stack and Gundlach (1992) conducted a study of the relationship between country music and suicide. Analyzing data from 49 metropolitan areas, they found a significant effect of country music air time on white suicide rates, controlling for southern region, poverty, divorce, and gun availability. In a subsequent replication attempt (Maguire & Snipes 1994), we could not reproduce this effect, finding instead no multivariate association between the two constructs. Stack and Gundlach (1994) responded with a criticism of our methods. The purpose of this article is to readdress the issue, showing that — at least in Stack and Gundlach's sample of metropolitan areas — country music air time is unrelated to white suicide rates.

Stack and Gundlach have responded to our critique of their study on country music and suicide in an understandable fashion. They side-step our fundamental criticism (that their country music data does not come from the source it is supposed to have come from) and address an issue that is irrelevant (the definition of metropolitan areas). In their original article, Stack and

* *Both authors contributed equally to this article. Direct correspondence to Jeffrey Snipes, School of Criminology and Criminal Justice, Florida State University, Tallahassee, FL 32306-2025.*

 Social Forces, September 1995, 74(1):327-329

TABLE 1: Relationship of Country Music to Suicide

	Beta
Controls	
None	.46***
Divorce	.24
Divorce and poverty	.15
Divorce, poverty, and southern region	.08
Divorce, poverty, southern region, and gun availability	.05
(N = 48)	

[a] In our original critique we used all fifty metropolitan areas for which country music were available. However, the WSR data provided by Stack and Gundlach contain data for only 48 of these locations.

*** $p < .001$. None of the other betas were significant at the .05 level.

Gundlach reported that their country music came from the 1985 *Radio and Records Rating Report*. We obtained this data from the source, and used it in all our analyses. Although this was the data listed in their references, we subsequently learned from them that *they did not use this data.* This comports well with our original suspicion that "the discrepancy may be based on a misrecording of the country music data" (Maguire & Snipes 1994).

In their rejoinder to our critique, Stack and Gundlach employ four measures of country music: (1) data from Maguire and Snipes, which is the exact data listed in their original article as *their* source, (2) *Radio and Records Rating Report* 1986, (3) Stack and Gundlach data, for which no source is listed, and (4) mean of 1985-86 data. The use of all these data sources is puzzling, especially since they did not really use the data listed in their original references. We have no indication of the source for the Stack and Gundlach country music data which they discuss in their rejoinder.

They address this issue only indirectly by arguing that "our" country music data is highly correlated with theirs. This does not, of course, mean that the various measures are interchangeable. For example, their country music data is correlated with southern region (a binary variable with *no* controversy in its measure) with a Pearson r of .26. The correlation between our country music and southern region, on the other hand, is .60.[1] Given that their primary independent variable has been inaccurately recorded, resulting in substantially

inaccurate correlations between it and other independent variables, it seems unnecessary to address measurement error in the dependent variable.

Stack and Gundlach argue that the contradictory findings between the studies is due to the discrepancy in the suicide rates. To show that this is not true, we simply substituted their suicide rates for ours and reran the model. After the substitution, the effect of country music remained insignificant. We found the same bivariate correlation between country music and suicide that Stack and Gundlach report in their rejoinder ($r = .46$); thus these data are not an issue in this analysis. Our results confirm our initial suspicion that there is no direct relationship between country music and suicide. Rather, as Table 1 shows, the relationship is spurious, attributable primarily to the effects of divorce and poverty, and to a lesser extent southernness and gun availability.

This is an *exact reproduction* using the *exact same data sources* of the model which Stack and Gundlach reported in their initial article. Because they questioned our critique on the basis of our white suicide rates, we simply replaced our rates with theirs (which they sent to us) and reran the model. They also pointed out that our control variables will be flawed to the extent that the populations in their denominators are not matched with the mortality areas. To correct for this, we took the area definitions sent to us by Stack and Gundlach, and calculated populations, divorce rates, and poverty rates, according to these mappings. As shown in Table 1, the effect of country music becomes insignificant when we introduce controls into the model.

Although we cannot speculate on why Stack and Gundlach, in their rejoinder, did not replace their country music data with ours (with which we provided them) and rerun the full multivariate model, it seems that this would have been the most convincing method of showing our original analysis to be faulty. Regardless of the ambiguity in *why* our analysis fails to replicate Stack and Gundlach's findings, this exchange has been important in showing how the smallest fluctuations in data measurement and analysis techniques can affect the conclusions of studies using small samples at the aggregate level.

Notes

1. This is for the 50 cities in our sample. For the 48 cities which are in both our data set and Stack and Gundlach's, this correlation is .57.

References

Radio and Records. 1985. *Radio and Records Rating Report and Directory*. Los Angeles: Radio & Records. Spring.

Maguire, Edward R., and Jeffrey B. Snipes. 1994. "Re-Assessing the Link between Country Music and Suicide." *Social Forces* 72:1239-43.

Stack, Steven, and Jim Gundlach. 1992. "The Effect of Country Music on Suicide." *Social Forces* 72:211-18.

____. 1994. "Country Music and Suicide: Another Look." *Social Forces* 72:1245-48.

Country Music and Suicide — Individual, Indirect, and Interaction Effects: A Reply to Snipes and Maguire

STEVEN STACK, *Wayne State University*
JAMES GUNDLACH, *Auburn University*

Abstract

Previous work on country music and suicide has neglected individual-level data analysis of country music fans' suicide risk, use of an average annual country music exposure index, an analysis of indirect effects, and an assessment of interaction effects. The present article addresses these issues. First, country music fans are found to be at significantly higher risk of suicide than nonfans both in gun ownership and marital disruption. A significant zero-order relationship between country music and suicide does become spurious with controls when we use an average annual exposure index. However, country music exerts both indirect and interaction effects on suicide. Hence, the present article offers additional support for a link between country music and suicide.

First, we would like to employ some recently released, microlevel data that has relevance to the theory relating country to suicide. The debate over the country music subculture's impact on suicide had not used such individual level data to date. While content analysis of country music songs often suggests that country music fans may have elevated suicide risk to the extent that their life styles mirror the life styles found in the songs, there has been a dearth of empirical work on this central issue. For example, we found associations between country music and divorce, a known risk factor for suicide (e.g., Stack and Wasserman 1993). However, we had no microlevel data documenting that country fans were more apt to experience marital disruption than nonfans. Data are needed to document this association at the micro level of analysis.

Recent data from the General Social Surveys can be used to address this key issue (Davis 1993). In 1993 for the first time the GSS polled persons' musical preference. Some caution needs to be exercised in interpreting the results from these data since they are for 1993. By 1993 country's audience had grown and, according to our critics, had become more middle class in its orientation. Any

* *Direct correspondence to Department of Criminal Justice, Wayne State University, Detroit, MI 48202.*

relationship between country and suicide risk factors may have disappeared, or become weaker, as a result, by 1993. Herein a country fan is defined as someone who responded that they like country music "very much." There were 248 such country fans and 632 nonfans in the GSS sample in 1993. Suicide risk factors are divorce and gun ownership. Divorce is defined as a positive response to the question: "Have you ever been divorced or separated?" Gun ownership is defined as the subject's reporting that they have a gun in their home.

Table 1 presents the data on divorce for country vs noncountry fans: 27.4% of the fans reported marital disruption as opposed to 18.4% of the nonfans ($\chi^2 = 8.85$, 1 df, $p < .003$). Country music fans are significantly more at risk of suicide on this suicide risk factor than nonfans.

Table 2 presents the data on gun ownership and musical preference: 61.8% of the country fans have a gun in their homes compared to 40.2% of the nonfans ($\chi^2 = 32.69$, 1 df, $p < .000001$). Country fans are at a significantly higher risk level for suicide compared to nonfans on this factor as well.

Given that country music fans are at higher risk of divorce and gun ownership than nonfans, we argue that country fans are also at higher risk of suicide.

Second, there is some confusion over the Radio and Records data on country music. Radio and Records publishes its *Rating Report and Directory* between two and three times annually. Estimates of country market shares are available for at least the spring and fall of each year. Snipes and Maguire (1995) seem to have the impression that just one estimate is available for each year. This is not the case. There are problems, however, in obtaining these data including the fact that the publication of these estimates is staggered (e.g., the Fall ratings for a given year come out in the next calendar year). Also, few libraries subscribe to the series and those that do typically discard old issues such as those giving 1985 estimates.

Probably the best measure of country's market share for 1985 is an average of the Spring and the Fall 1985 market shares data. We did not do this in our article (Stack & Gundlach 1992). We now have all the data (Radio and Records 1985; 1986). Taking the mean country market share for Spring-Fall 1985 as the measure of the strength of the country music subculture, the correlation between country and urban suicide rates is $r = .48$ ($p < .05$). This compares to an $r = .54$ in Stack and Gundlach (1992). While the difference in r's appears small, the relationship using the mean 1985 country market share could be spurious, as Snipes and Maguire (1995) correctly caution.

To address the issue of possible spuriousness, we replicated our model substituting the mean annual country music market share for 1985 as the measure of country music. We find that the significant zero-order relationship between country music and white suicide is no longer significant in this model.

Third, the idea, however, that country music has no effect on white urban suicide may be premature. Our model is overly conservative given that it focused only on direct effects. It is possible that country might be associated with suicide through indirect effects, as well as interaction effects.

Country music may have an indirect effect through its association with one or more of the predictor variables in the model. For example, given its preoccupation with the travails of love, it is conceivable that country might be

TABLE 1: The Relationship between Musical Preference and Ever Having Been Divorced or Separated, 1993

Divorced or separated	Country Fan (Percent)	Noncountry Fan (Percent)
Yes	27.4	18.4
No	72.6	81.6

$\chi^2 = 8.85$
$p = .00293$

predictive of urban divorce rates. Three fourth's of country hit songs dwell on themes such as marital strife and dissolution (Rogers 1989). This is important since divorce and suicide tend to be strongly related (e.g., Lester 1992). Taking the urban divorce rate as a dependent variable, we incorporated the remaining predictor variables as X variables affecting divorce. Country music had the strongest relationship to divorce ($\beta = .42$). Divorce, in turn, is the strongest predictor of suicide ($\beta = .39$). Hence, the indirect effect of country on suicide amounts to .42 x .39 or .16. Through its concern with marital dissolution and strife, it is conceivable that country music fosters divorce which, in turn, increases the risk of suicide.

Country music might also influence suicide through interaction effects with the other predictor variables. For example, in order for country music to have an impact on the urban suicide rate, perhaps it must be combined with an audience that is predisposed to suicide. A good candidate for such an audience would be the divorced population. That is, it would be in cities with both a strong country music subculture and a relatively large divorced population that we would anticipate a high suicide rate. Further, from the standpoint of a differential identification theory of deviance (e.g., Matthews 1968), it would be anticipated that divorced people would be precisely one of the most likely groups to differentially identify with the travails of love theme in country music. Divorced people would be the most apt to be receptive to the sad messages in such songs and be at greater suicide risk. For a further discussion of differential identification and suicide risk see Stack (1987).

To test a nonadditive model of suicide, divorce, and country music, we reestimated the model with a multiplicative term, divorce x country music. We used the mean level of country music for the year 1985 as the country music subculture estimate. Given multicollinearity between the multiplicative term and its components, only the multiplicative term was included in the equation, along with the measures of gun availability, southern location, and the structural poverty index. Controlling for the other variables in the equation, the divorce x country term was significant ($b = 0.02$, $t = 1.87$, $p < 0.05$). Cities that combine a high level of divorce with a high level of country music have an elevated suicide rate.

TABLE 2: The Relationship between Musical Preference and Having a Gun in the Home, 1993

	Country Fan (Percent)	Noncountry Fan (Percent)
Having a gun in the home		
Yes	61.8	40.2
No	38.2	59.8

$\chi^2 = 32.69$
$p = .00000$

Given that country music has both indirect and interaction effects that influence suicide, it is not correct to conclude that it is irrelevant to the generation of urban suicide rates. There are additional reasons for not dismissing the country music subculture from the etiology of suicide, and some further cautions as well.

Fourth, the relationship between country music and suicide needs replication for at least two reasons. The estimates of country music radio market shares fluctuate significantly enough from survey to survey to warrant analyses for other years in order to get more reliable results. Second, it is possible that given the increased popularity of country music over the last ten years, there is no longer even a zero-order association between it and suicide by the mid-1990s. What we found for 1985 may be a historically specific phenomenon. As the audience of country music shifts to a more middle class group with a lower divorce rate and generalized receptivity to the suicidal mood found in many classic country hit songs, it is anticipated that such songs would have less of an impact on such an audience. Indeed, such an audience could significantly change the content of country songs such that there might be fewer sad songs on the travails of love. These arguments are speculative and await a content analysis of country hit songs similar to the work of Rogers (1989) on 1,400 such songs from an earlier period.

Alternative measures of the country music subculture are needed in order to systematically explore the possible link between country music and suicide. For example country music magazine subscriptions could be used to this end, periodical circulation being another measure of culture (e.g., Baron & Straus 1989; Stack et al. 1994).

To the extent that the country music is stronger in rural areas, the study needs further replication. Our investigation was restricted to large urban centers. Perhaps the country music subculture would have a stronger impact on suicide where it is more dominant, that is, in rural areas.

Finally, it is also possible that the country music subculture fosters alcohol consumption. Lyrics in country often portray drinking as a normative response to stress in life (Chalfant & Beckley 1977). If so, that country may have still another indirect effect on suicide. An analysis that incorporates an index of

alcohol consumption is needed since alcohol consumption has a direct bearing on suicide risk (e.g., Stack & Wasserman 1993).

In summary, the analysis of individual level data connects being a country fan to known suicide risk factors, country has strong indirect affects on suicide, and country exhibits an interaction effect with divorce on suicide. It seems that the critics' notion of spuriousness in the country and suicide relationship is flawed.

Generally speaking, work is needed to explore the relationship between art and suicide, including country music. Elsewhere a relationship between heavy metal music subculture and youth suicide has now been documented (Stack, Gundlach & Reeves 1994). Possibly the rap music subculture might foster a high incidence of black youth suicide in urban areas.

References

Baron, Larry, and Murray Straus. 1989. *Four Theories of Rape in American Society: A State Level Analysis*. Yale University Press.

Chalfant, Paul, and Robert Beckley. 1977. "Beguiling and Betraying: The Image of Alcohol Use in Country Music." *Journal of Studies on Alcohol* 38:1428-33.

Davis, James. 1993. *General Social Surveys, 1972-1993: Cumulative Codebook*. Storrs: Roper Center.

Lester, David. 1992. *Why People Kill Themselves*. 3d ed. Charles Thomas.

Matthews, Edward. 1968. "Differential Identification: An Empirical Note." *Social Problems* 15:376-83.

Radio and Records. 1985. *Radio and Records Rating Report and Directory*. Los Angeles: Radio & Records. Spring.

Radio and Records. 1986. *Radio and Records Rating Report and Directory*. Los Angeles: Radio & Records. Fall.

Rogers, Jimmie N. 1989. *The Country Music Message*. University of Arkansas Press.

Snipes, Jeffrey, and Edward Maguire. 1995. "Country Music, Suicide, and Spuriousness." *Social Forces* 74:327-29.

Stack, Steven. 1987. "Celebrities and Suicide: A Taxonomy and Analysis." *American Sociological Review* 52:401-12.

Stack, Steven, and Jim Gundlach. 1992. "The Effect of Country Music on Suicide." *Social Forces* 71:211-18.

Stack, Steven, James Gundlach, and Jimmie L. Reeves. 1994. "The Heavy Metal Subculture and Suicide." *Suicide and Life Threatening Behavior* 24:15-23.

Stack, Steven, and Ira Wasserman. 1993. "Marital Status, Alcohol Consumption and Suicide: An Analysis of National Data." *Journal of Marriage and the Family* 55:1018-24.

Book Reviews

The Bell Curve: Intelligence and Class Structure in American Life.
By Richard J. Herrnstein and Charles Murray. Free Press, 1994. xxvi + 845 pp. $30.00.

Reviewer: FRANÇOIS NIELSEN, *University of North Carolina at Chapel Hill*

This massive but well-written study by the late Harvard psychologist Richard Herrnstein and public policy analyst Charles Murray is about the role of intelligence in the social structure of American society. Their discussion is organized into four parts. In part 1, they piece together a picture of the emergence of a cognitive elite in American society in the course of this century, through the increasing concentration of individuals with high IQ into the ranks of the educated elites and into high prestige occupations. This trend is driven by the reduced importance of social origin as the major factor of class distinctions and the increasing value of intelligence in the marketplace of modern industrial society, due to the increasing scale of organizations and markets combined with the intellectual demands of complex technologies and intricate regulatory structures. The authors view the cognitive elite as becoming increasingly separated from the rest of the population by a widening economic gap and increasing physical separation in the workplace and in residential communities. Through assortative mating, this process of "cognitive partitioning" is viewed as leading to the creation of a quasi-hereditary cognitive elite. This macrosociological description constitutes an updated and more richly documented version of the earlier work by Herrnstein (1973) on IQ in a meritocracy; sociologists will also find this discussion reminiscent, in a less fictional register, of the work of Michael Young (1958).

In part 2, Herrnstein and Murray present their own empirical analyses of data from the National Longitudinal Survey of Youth (NLSY), based on a longitudinal survey of about 12,000 youths aged 14 to 22 in 1979. Their analytical strategy is to estimate logistic regressions of a wide variety of dichotomous socioeconomic outcomes (dropping out of school, graduating from college, being in poverty, being unemployed, committing a crime, having a child out of wedlock, etc.), showing that in most cases intelligence, measured as the score on the Armed Forces Qualification Test (AFQT), has a stronger effect on the outcome than a composite measure of SES (based on family income and parents' education and occupational prestige). These analyses are carried out with non-Hispanic whites only, so issues of race or ethnicity are not addressed at this stage of the analysis.

Part 3 introduces issues of race and ethnic differences in IQ. The authors review the large literature showing the average IQ of blacks as being about one standard deviation lower that the white average, and the East Asian average (in the U.S. and abroad) as several points higher than the white mean. While maintaining a position of official agnosticism with respect to a possible genetic basis for these group differences in IQ, the authors suggest that the detailed pattern of the differences —

such as the systematic difference in intellectual profile between East Asians and whites, with an East Asian advantage concentrated in the nonverbal intelligence scores, and the finding that black and white scores differ most on the most *g*-loaded tests (i.e., those most related to a general intelligence factor) — strongly points to genetic roots. They go on to estimate the effect of IQ differences on average group differences in a variety of socioeconomic outcomes in the NLSY data. For many socioeconomic outcomes, such as family income and proportion below the poverty line, controlling for IQ substantially reduces the gap between blacks and whites. In some cases, such as wages, the discrepancy vanishes almost entirely when IQ is controlled. For educational attainment and access to professional occupations, however, the gap is reversed so that controlling for IQ, blacks have higher educational and occupational achievement than whites. This part of the book also discusses dysgenic effects of the higher rate of births among people of lower cognitive ability and the apparently contradictory phenomenon of the rise in average IQ that took place in industrial societies during the past few decades (the Flynn effect) and documents from the NLSY data the higher prevalence of low cognitive ability among people visited by a variety of social problems from dropping out of school to ever being interviewed in jail.

Part 4 deals with topics variously related to policy. The authors document the failure of many attempts to raise IQ. While better nutrition may play a role in the historical upward trend in IQ, formal schooling and special programs, notably Head Start, do not seem to raise IQ. Some highly intensive (and therefore costly) programs directed at young children, as well as adoption, might produce an IQ improvement. The authors credit affirmative action in higher education for opening college opportunities for more blacks and Latinos than would have otherwise been there, but spell out the downside of affirmative action programs in producing unfairness, distrust in the larger society of the achievement of minority students, and lowered productivity because of the suboptimal matching of ability and occupation. With respect to affirmative action in the workplace, they call for "scrapping the existing edifice of job discrimination law." In the penultimate chapter, the authors amplify their pessimistic view of a future "custodial state" in which an isolated cognitive elite (now merged with the affluent) protects itself in gated communities from the underclass at the bottom of the cognitive ability distribution, whose leadership has been drained and whose quality of life deteriorates further in the urban ghettos in which it is concentrated. The last chapter of *The Bell Curve* offers the authors' hopeful vision of an alternative society based on a return to the central role of local communities in which neighbors take care of each other, rules of conduct are made simple to understand and to follow, and everyone finds a "valued place" irrespective of their ability. This vision of vibrant local communities is hardly sinister, but the authors do not say how the devolution to local structures would happen, or even what trend might possibly lead in that direction. This exercise in utopia is largely independent of the rest of the book.

The Bell Curve has acquired its aura of controversy by challenging deeply held beliefs that make up the ethical-political credo of a large segment of the media and the public, including a large proportion of the sociological profession. While many people accept intuitively, at least in private, the notion that there is such a thing as intelligence, and that people differ in how intelligent or "smart" they are, the notion that part of the difference in intelligence among individuals might have a genetic basis is harder to accept, as it seems to imply a greater inevitability to the inequality in the distribution of ability than if cognitive ability were determined by purely

environmental factors that can be changed (it is hoped) through deliberate policy intervention. The idea that average intelligence varies among racial and ethnic groups, and that part of this variation might be genetic is the most scandalous of all. Although the authors do not claim to be certain about a genetic basis of group differences, they review the compelling circumstantial case for the existence of such a genetic basis, thereby committing the ultimate blasphemy against the dominant egalitarian social philosophy of our times. Many sociologists who already find such ideas disturbing as ordinary citizens may be additionally predisposed against attributing a central role to intelligence in explaining social organization by the preposterous notion that social facts should only be explained in terms of other social facts, so that intelligence — being a psychological, not a social, concept — does not constitute a legitimate sociological explanation of social reality.

Amidst the buzz and gossip that greeted the publication of the book in the fall of 1994, there was an almost palpable expectation of a critique that would deal the book a decisive blow by uncovering the "smoking gun," the irreparable defect that would invalidate the book as a whole and thereby save the conventional orthodoxy. Even highly competent reviewers in the social sciences have not discovered this fatal weakness, some contributing valid criticisms (e.g., Goldberger & Manski n.d.; Heckman 1995), others expressing their disagreement with little more than invectives (Hauser 1995). The task of hostile critics is made difficult by the fact that much of *The Bell Curve* consists of reviews of existing literatures, including the massive literatures on behavioral genetics and the psychometric properties of IQ tests. The authors' interpretation of these bodies of work is by no means outlandish: their summation corresponds by and large to a cautious and even conservative interpretation of the main thrust of the literature, for example in their working estimate of the heritability of IQ in the white population at .6, which is on the low side relative to current estimates in behavioral genetics (Scarr 1994-95). Many of the authors' conclusions such as those concerning the unbiasedness of IQ tests, the heritability of IQ, and the meaning of race and ethnic differences are widely accepted findings within the specialized literatures in which they originated. They appear novel and controversial only because the bulk of this research has been carried out away from the public eye and from the attention of many social scientists. The results of this research have been hidden in specialized journals, to emerge to collective consciousness only during brief episodes of intense public debate of the kind that greeted the work of Jensen (1969) or Herrnstein (1973). With scattered exceptions (e.g., Eckland 1967; Gordon & Rudert 1979; Scarr & Weinberg 1978), these findings have not been well disseminated in the sociological literature. It has been hard for critics of *The Bell Curve* to challenge convincingly the parts of the book that report well-established findings of sophisticated bodies of research.

The other main source of evidence used by the authors is their own analysis of the NLSY data. In a typical analysis, they estimate the coefficients of the logistic regression of a dichotomous outcome, such as dropping out of high school, on IQ, SES, and age (as a control for the dependence of the AFQT measure of IQ on age). They present the results of each regression graphically, by contrasting on the same graph the curve of the estimated probability of the outcome as IQ ranges from -2 to +2 standard deviations, with SES and age kept constant at their mean sample values, with the curve of the estimated probability of the outcome as SES ranges from -2 to +2 standard deviations, with IQ and age kept constant at their mean sample values. They show that, for many outcomes, the curve representing the probability of the outcome as a function of IQ is steeper than the corresponding curve for SES, and

conclude that IQ is a more important factor than SES in determining the outcome. As noted by some reviewers (e.g., Goldberger & Manski n.d.), this effective graphical presentation is equivalent to the use of standardized coefficients to compare the relative effects of independent variables measured in different metrics. While many sociologists would go on from here to estimate more complex models including larger sets of independent variables, the simple analysis serves well to reinforce the authors' central contention that IQ is a vastly better predictor of important aspects of social performance than SES, a variable that is the centerpiece of so many sociological explanations of such outcomes.

The larger scientific context in which the authors present their empirical findings is unusual, compared to many previous debates in the social sciences, in that the data they use are so publicly accessible. A CD ROM with the NLSY data can be obtained for a small sum, and many readers of this review can easily replicate the authors' analysis on their own personal computers. Social science research with controversial implications has rarely been as open and aboveboard as this, and one expects that many replications and extensions of the authors' analysis will appear in the future. The authors' analysis has been criticized on the grounds that the composite SES measure is unlikely to capture the entire childhood experience of subjects (which is a criticism of the concept of SES in general, not only the authors' use of it), and that their regression models are underspecified, in particular by not including education as an independent variable to predict other outcomes, since education could affect these outcomes independently of IQ (Heckman 1995). Some critics have also noted that the pseudo-R^2s that measure the fit of the logistic models are low for some outcomes. (The measure of fit is substantially higher for educational outcomes, such as dropping out of high school or graduating from college, and for some socioeconomic outcomes such as welfare dependency, than for the other outcomes analyzed by the authors.) Here also, critics are unlikely to find a smoking gun, because the strong effect of IQ that the authors find in their simple models is quite robust, as researchers familiar with the NLSY data know. No amount of nitpicking at the details of the authors' analysis is likely to obscure the remarkable fact that IQ, a measure of a psychological construct, is the best predictor of many aspects of the social performance of individuals many years later, with an effect stronger than any other of the traditional "background" variables including SES, certainly one of the sacred cows of sociological research.

While the major empirical findings of the authors are unlikely to be overturned, the book is not without weaknesses. As pointed out by Heckman (1995), the 875 pages of the book do not constitute a single coherent thread of argumentation, as some issues that are discussed at length by the authors, such as the question of the heritability of intelligence, are arguably separable from the rest of the issues; and the policy vision, particularly the rather romantic account of life in strong local communities, does not follow in any direct fashion from the evidence of the previous chapters. Occasional interpretations of current theories of behavioral genetics are incorrect, for example the discussion of the consequences of reducing the variability of the environment (p. 106), and other passages in which the authors exaggerate the intergenerational rigidity of cognitive classes based on genetic transmission of intelligence. Curiously, the same tendency to exaggerate the rigidity of genetic mechanisms — by downplaying the importance of regression to the mean — had already been criticized in the earlier work of Herrnstein (1973) by Eysenck (1973), who showed that genetic transmission entails much greater intergenerational mobility among IQ categories than Herrnstein had acknowledged.

A major contribution of *The Bell Curve* to sociological research is its polished presentation of the findings of literatures that have been by and large hidden from social scientists, including the behavioral genetics literature on the strong heritability of IQ, the conclusions of psychometric research on the absence of race (but not linguistic) bias in major IQ tests (Jensen 1980), the studies in the field of human resources of the predictive validity of IQ with respect to job performance, and the program evaluation literature revealing the failure of many intervention programs aimed at raising the cognitive ability of children. This skillful review, together with the wide diffusion of the book and the public attention that it has generated may help unravel the veil of collective denial that has shrouded knowledge of these issues in much of our field. With respect to the genetic basis of individual differences in cognitive ability, the book must also be viewed as part of the larger movement of ideas toward the rehabilitation of biological explanations and evolutionary approaches in the social sciences that has been discussed by Degler (1991). This emerging movement is associated with the discovery of a biological basis for many kinds of behaviors (including gender roles, Udry 1994), as well as the ascent of sociobiology (Nielsen 1994).

Sociology has a natural place to occupy within the scientific division of labor that would tackle the issues raised by Herrnstein and Murray, which is to focus on the aggregate consequences of the central role of IQ for individual outcomes demonstrated by them. Their analysis of processes of cognitive partitioning in modern industrial society in part 1 of the book is evocative but lacks formal precision, as it is pieced together from partial data and case studies with much extrapolation. There is a clear need to better document the existence of the mechanisms of social selection on the basis of cognitive ability postulated by the authors and integrate them with theories of social stratification. A special study under such a heading would include the possible role of cognitive partitioning on the upswing in income inequality that has taken place over the last 25 years. The possible dysgenic consequences of the negative correlation between fertility and IQ are another natural topic of study that is already being investigated by sociologists (Coleman 1993; Preston & Campbell 1993). Finally, in much sociological research on educational, occupational, economic, and behavioral outcomes of individuals, it will be harder than ever for researchers to ignore the importance of IQ as a potential explanatory factor with lame excuses about intelligence being a "psychological variable" and therefore off limits for sociologists.

References

Coleman, James S. 1993. "Comment on Preston and Campbell's 'Differential Fertility and the Distribution of Traits'." *American Journal of Sociology* 98:1020-32.

Degler, Carl N. 1991. *In Search of Human Nature: The Decline and Revival of Darwinism in American Social Thought*. Oxford University Press.

Eckland, Bruce E. 1967. "Genetics and Sociology: A Reconsideration." *American Sociological Review* 32:173-94.

Eysenck, Hans J. 1973. *The Inequality of Man*. Edits.

Goldberger, Arthur S. and Charles F. Manski. Forthcoming. "Review Article: *The Bell Curve* by Herrnstein and Murray." *Journal of Economic Literature.*

Gordon, Robert A., and Eileen Rudert. 1979. "Bad News Concerning IQ Tests." *Sociology of Education* 52:174-90.

Hauser, Robert M. 1995. "Review of *The Bell Curve." Contemporary Sociology* 24:149-53.

Heckman, James J. 1995. "Cracked Bell. (Review of *The Bell Curve.*)" *Reason* 26(March):49-56.

Herrnstein, Richard J. 1973. *IQ in the Meritocracy.* Atlantic-Little Brown.

Jensen, Arthur R. 1969. "How Much Can We Boost IQ and Scholastic Achievement?" *Harvard Educational Review* 39:1-123.

____. 1980. *Bias in Mental Testing.* Free Press.

Nielsen, François. 1994. "Sociobiology and Sociology." *Annual Review of Sociology* 20:267-303.

Preston, Samuel H., and Cameron Campbell. 1993. "Differential Fertility and the Distribution of Traits: The Case of IQ." *American Journal of Sociology* 98:997-1019.

Scarr, Sandra. 1994-95. "What Is Equality? (Review of *The Bell Curve.*)" *Issues in Science and Technology* 11:82-85.

Scarr, Sandra and Richard A. Weinberg. 1978. "The Influence of 'Family Background' on Intellectual Attainment." *American Sociological Review* 43:674-92.

Udry, Richard J. 1994. "The Nature of Gender." *Demography* 31:561-73.

Young, Michael. [1958] 1961. *The Rise of the Meritocracy, 1870-2033: An Essay on Education and Equality.* Penguin Books.

Farm Security Administration Photographs of Florida.
Edited with an original essay by Michael Carlebach and Eugene F. Provenzo Jr. University Press of Florida, 1993. 124 pp. Cloth, $39.95.

Reviewer: AUGUSTUS BURNS, *University of Florida*

In no area of American society did Franklin Roosevelt's New Deal more radically alter the political and economic landscape than in the nation's farm life. The myriad programs that Congress and the President enacted, as Clifton Luttrell has remarked, "represented a new policy of government intervention in the business affairs of individual farmers. They marked a turning point from a free to a highly controlled economy." And the consequences of that policy shift have been the subject of a persistent and intense debate from the moment those policies were initiated.

The debate was joined in part by the very governmental agencies charged by Congress and the President to restructure American agriculture. Eager to persuade American citizens that farmers and farm workers were severely distressed — an easy case to make in the Depression years — the New Deal bureaus became their own advocates. In one such effort, farm policy administrators went into the picture business.

In 1935, President Roosevelt issued an executive order that created the Resettlement Administration, an agency intended to develop programs to assist farmers and their families who had been dispossessed of their lands (some of which dispossession had been accelerated by other New Deal farm programs, most notably the first Agricultural Adjustment Act).

To head this agency, Roosevelt named Rexford Tugwell, a Columbia University economics professor before becoming a member of FDR's New Deal "Brains Trust." Within a few months of beginning his duties, Tugwell had managed to offend nearly every segment of American agriculture, becoming a target of ceaseless criticism. In response, Tugwell launched a counter-attack, seeking to mute his critics by making a case for his policies directly to Congress and to the American people. He thus

employed Roy Emerson Stryker, formerly a Columbia University colleague to Tugwell, to assemble a group of photographers. Their mission would be, on government payroll, to document in pictures the plight of American farmers and farm workers — the agriculturally dispossessed. "Our job," Stryker noted in 1951, "was to educate the city dwellers to the needs of the rural population."

Stryker brought on board a team of genuinely gifted photographers: Carl Myradans, Walker Evans, Ben Shahn, Dorothea Lange, John Collier, Arthur Rothstein, Marion Post Wolcott, John Vachon, and Gordon Parks, a group that could form a nucleus for an American photography Hall of Fame. Predictably, they produced a photographic archive that was simply stunning in its technical quality and in its artistic construction.

One of the states where these photographers did much of their work was Florida. Much like California, Florida was a state where many people displaced from their farms in other parts of the South came to search for work and survival. They came in part because the depression years were a period of tremendous upheaval in Florida agriculture. The industry was essentially revolutionized into the agribusiness complex that became a vital part of the "new Florida." In the thirties, new farms of very large acreage were opened up in areas of the state previously little used for farming, most notably the area south of Lake Okechobee. Here up to 50,000 migratory workers might find work in peak season, often living in unimaginable squalor. Central Florida agriculture likewise grew substantially in the thirties.

The observable result of this boom in the midst of depression was a contrast of economic conditions that seemed almost medieval in its gross disparities. For Florida remained a tourist mecca to those people fortunate enough to be able to afford winter vacations from colder northern climates; and the operators who ran the large farms and processing plants survived the depression years in relative comfort.

In sum, the contrasts among the people in Florida was a socially conscious photographer's dream. Documenting rural poverty in Florida was a cinch, and the contrast with the manner of life in Palm Beach or other tourist communities made the economic differences even more striking.

In their superbly crafted work *Farm Security Administration Photographs of Florida*, Michael Carlebach and Eugene F. Provenzo Jr. have produced a book that faithfully reflects both the achievement and the craftmanship of the talented men and women who made these remarkable photographs. They have, in addition, produced a work that details in arresting visual images a "rural world lost," to borrow Jack Temple Kirby's phrase. Florida before World War II, before Disney, and all the other economic and political changes that have overtaken this semi-tropical sand bar, was indeed a different place. The ways in which it was different are too numerous to examine in a brief review. But as these photographs make clear, Florida was both more beautiful and more terrible, more promising and more forbidding. Clearly, the homogenization of space and structure that Daniel Boorstin identified many years ago as a product of modern America has nowhere occurred with greater impact than in Florida. To examine these pictures and to contrast them with modern Florida is to conclude that Joseph Schumpeter's phrase for industrial transformation — "creative destruction" — is a woeful understatement when applied to the "Sunshine State."

Carleback and Provenzo introduce their work with a very insightful essay that places these photographs in their proper historical context. They trace the evolution of Florida agriculture and the political circumstance out of which the Resettlement Administration (reorganized as the Farm Security Administration) came into being. They make clear that the photographs were in a sense essentially a political undertaking, intended to be used by the Roosevelt Administration to advance its

own farm agenda. And the discussion of these photographs is thus sensitive to the ongoing political disputes that accompanied them.

As pieces in a bitter political contest, these photographs have lost most of their partisan tone. The years since the New Deal have produced such extensive debate regarding the effect of New Deal farm policy that the photographs now are enveloped in that ambiguous legacy. What remains, the authors/editors write, is a historical and cultural artifact that also happens to be, collectively, a work of art. Both editors and the University Press of Florida are to be commended for the critical and production excellence that have combined to produce this exemplary volume.

Conditions of Liberty: Civil Society and Its Rivals.
By Ernest Gellner. Allen Lane, Penguin Press, 1994. Cloth, $27.95.

Reviewer: CHARLES KURZMAN, *Georgia State University*

The hypotheses in this book are thought-provoking, wide-ranging, and almost entirely unproven. Ernest Gellner compares four sociopolitical ideal types: premodern segmented society, civil society, Communist society, and Islamic society. He seeks to understand just how these forms differ, in their essences, and how they came to differ. This is a tall order for a short book, which gives the impression — a chatty manner and relative lack of documentation — of arising from a set of lectures. But Gellner's long experience as a philosopher and social researcher makes this a provocative new take on big thoughts in social theory.

Few theoreticians of the West's Great Transformation can draw on Ibn Khaldun's analysis of urban commercial society and tribal militarism. Few could make the analogy between Communist eschatalogy and the Islamic concept of Umma. Gellner's familiarity with Eastern Europe, both before and after 1989, allows him to note the widespread Soviet cynicism in the 1970s and 1980s about central tenets of Marxism: for instance, the "wry smile" of Soviet scholars passing a prominent red banner that quoted Lenin to the effect that "the teaching of Marx was all-powerful because it was true."

Gellner's argument, in short, is that civil society requires political centralization and socioeconomic decentralization. The centralization involves the building of nation-states with a monopoly on coercion — which Weber took as the definition of the state, but which in fact took centuries to accomplish. Only when this monopoly has been constructed, Gellner says, can segmented society — the localism of fiefdoms, tribes, and city-states — be overcome.

However, the creation of a nation-state does not necessarily lead to the flourishing of civil society. It can as easily lead to a Communist or Islamic social order. Islamic society, Gellner says, sacralizes everyday life, and subsumes economic to moral ideals. Gellner contrasts this image of the Umma — the community of believers, which punishes apostates with death, at least in theory — and a concept of "modular" individuals, analogous to modular furniture units, who combine and recombine easily into various agglomerations. Modern individuals do not need sacrificial rituals to mark their entry into a political party, and they are not accused of apostasy or treason for leaving. Although these agglomerations are not all encompassing, they are powerful enough to be binding on their members for short periods and discrete purposes.

Communism also undermines civil society by mandating a "secular Umma," in Gellner's phrase, in which public control of the means of production is intended to usher in a new form of liberated human relations. It failed not because of Stalinist excesses or Brezhnevian mismanagement — though the squalor of the Brezhnev

years hurt it more than the terror of the Stalin years — but because state control of the economy is necessarily fatal to civil society. Gellner makes the provocative argument that not only is Marxism totalitarian, but "under modern conditions, any totalitarianism will also inevitably be Marxist." It may not reproduce or revere Marx's philosophy *per se*, but it will plow under the capitalist economy in order to eliminate rival centers of power.

The main problem with comparisons of ideal types is knowing just when and where they apply. Does the concept of Islamic society refer to the official imposition of Islamic law, as in Iran, Pakistan, Saudi Arabia, and elsewhere, or to countries with secular-nationalist regimes and Islamic oppositions, such as (at this writing) Algeria, Egypt, and Turkey? It's unclear, even, why they should be lumped together as a single sort of society. Yet Gellner's great-books approach to the subject leaves off Islamic history in the Middle Ages, with Ibn Khaldun. His occasional references to contemporary Iran and Morocco don't give much guidance as to the referent of the ideal type.

The Communist ideal type is somewhat more credible. Gellner wrote this book while helping to establish the Central European University in Prague, so he spends considerably more time discussing Communism than Islam. However, Gellner's antagonism — he admits to having been fooled, once, by Communism's promises of economic progress — is insufferable. Gellner takes the collapse of Eastern European Communism as final proof of the failure of Marxism, and spends an entire chapter arguing that Marxist critics of the Soviet Union were just as wrong-headed as Marxist supporters of the Soviet Union. They maintained the ideal of a non-capitalist industrial society even after it was proven in all sorts of contexts to have disastrous totalitarian results. "*All* possible combinations seem to have been tried out in the course of an exceptionally thorough controlled experiment kindly arranged by history, and the verdict seemed much the same: Marxist societies ranged from the repulsive to the unspeakable."

Gellner fails to notice that Marxism has never come to power in the advanced industrial countries that Marx predicted — which makes Marx a poor prophet, maybe even a poor social scientist, but hardly disproves the ideal. Indeed, Gellner avoids discussing any recent Marxist theory, much of which is in fact concerned with the need for civil society. The book makes much of the Eastern European rediscovery of the concept of civil society in the 1970s, calling this a natural reaction to Communist totalitarianism. But Western Marxists made a similar discovery around the same time: the destruction of the life-world by the administrative-bureaucratic apparatus and the rejuvenation of civil society came to occupy a central place in the postindustrial and new social-movement analyses of Jean Cohen, Alain Touraine, Jürgen Habermas, and many others. None of this literature is mentioned in the book.

Finally, the ideal type of civil society does not sound much like any particular country we know today. The features Gellner stresses include: modular citizens, who are members of groups but not too strongly attached to any one group; a sense of moral obligation that forces individuals to honor their commitments willingly, not because of social pressure; cultural homogeneity within a given nation-state, involving the elimination of "semi-private and local" languages and idioms; and rationality, including "the non-conflation of issues, the separating-out of the social strands." This is all just a bit too smug, and it ignores the threats to civil society which exist in the countries where is it supposed to be flourishing. To take just one issue: the rise of the professions in this century has been accomplished in large part through the creation of semi-private idioms, barriers to entry, and complex initiation rites. What has been the effect on modularity, homogeneity, and rationality?

And what is one to make of socialist democracies and other hybrids? Ideal types do not meld easily on a continuous spectrum, and Gellner is unclear on the topic. In successive paragraphs, Gellner ridicules hybrids — "Either productive units do have genuine autonomy (in which case they behave as in the market), or they do not" — and then defends what sounds like a middle position, with some state control: "The unfortunate consequence of the collapse of the Marxist Umma is that it seems to some people, quite incorrectly, to constitute a vindication of the complete marketization of society, a reinforcement of the minimal state and maximal market doctrine. This is most unfortunate." Elsewhere he calls the lack of an effective welfare state "unutterably repulsive." So Gellner is not a radical free-marketeer.

The book's image of civil society, however, is purely capitalist. Gellner dates the concept of civil society to Adam Ferguson in the eighteenth century but ignores the changes that the concept has undergone in the meantime. Ferguson's definition of civil society equates it with apolitical, self-interested, individual behavior, as opposed to the martial passions of politics. Commerce, not virtue, constitutes citizenship. By contrast, nineteenth century theorists from Tocqueville to Durkheim saw civil society as something far more complex and difficult. The definition shifted to the presence of intermediary groups between the individual and the state, which capitalism could not be counted on to generate automatically. These intermediary groups could not be apolitical and individualistic — they had to be group-oriented and politically involved precisely in order to counterbalance the power of the state. Purely self-interested individualism was now seen as a threat to liberty, and not its savior.

Gellner recognizes that intermediary groups are necessary for civil society: "how is it possible to have atomization, individualism, without a political emasculation of the atomized man (as in the world of Ibn Khaldun), and to have politically countervailing associations without these being stifling (as in the [segmented, premodern] world of Fustel de Coulanges)?" Gellner's answer, already noted, involves the concept of "modular" citizens, who can be mixed and matched in changing combinations within a given nation-state (though not between them). Capitalism involves modularity, Gellner suggests, and therefore has no difficulty creating and sustaining intermediary groups. The only societies in the modern world that have to worry about atomization are Communist ones.

This is fairly unsatisfying. Gellner avoids all the tough issues raised by Tocqueville, Durkheim, and others: why should capitalism imply civil society? In terms of Gellner's ideal types, couldn't capitalism just as easily strengthen nonmodular social organization? Gellner briefly discusses the cases of East Asian industrialization and notes how capitalism is accompanied there by near-feudal paternalism. (The same might be said of post-Soviet gangsterism.) The book also notes that Islam has become more powerful in the past century, as Islamic countries have been integrated into world capitalism. Finally, the dog that doesn't bark: mass society. Gellner dismisses the doomsday capitalist scenario of isolated mass individuals, abjectly subject to totalitarian rule, that has motivated so many theorists since the 19th century. Indeed, Ferguson may be seen as a forerunner of this obsession, since he worried about the ability of commercial individualists to ward off military despots. Gellner tells him not to worry. As states have grown stronger and technologically more advanced, the economy grew stronger as well. Ferguson's pessimism "came to be invalidated . . . by the tremendous expansion of productive power consequent on the impact of scientific technology." The image is almost one of a benign arms race between capitalism and the state.

It is hard to conceive of large corporations as defenders of liberty, helping to stave off totalitarianism. Even Gellner has difficulty maintaining this image for long. One of the fortunate aspects of British industrialization, he notes twice, was that it

occurred during a period of "a fairly feeble technology," powerful enough to counteract the state but "blessedly feeble enough not to destroy society or its environment, or give anyone power to dominate society militarily. All that has changed, and will never happen again. Modern technology is enormously powerful, and contains a disastrous potential for ecology and for the possibility of terrorism, by generating devastating weapons controllable by a small number of persons."

The state, therefore, must protect us from capitalist destructiveness — "In a sense, it is still true that 'we are all socialists now'" — just as capitalism protects us from state domination. Without appearing to notice the distinction, Gellner wishes to defend both types of liberalism: the nineteenth century version which promoted the free market against premodern restrictions, and the twentieth century version which promotes state regulation and redistribution.

There is nothing inherently wrong with this form of argument: the balance of power between capitalism and the state may indeed constitute the "conditions of liberty" that the title seeks. But it is not the argument that Gellner wishes to make. Despite his defense of the welfare state, he feels that state subordination is crucial to liberty. He states this clearly in his discussion of Western civil society, where "wealth leads to power, more than the other way around. This is both remarkable and exceptional. Marxism made it into a reproach when it should be a source of pride, and it [Marxism] absurdly generalized the dominance of production over an alleged political 'super-structure' as a law governing all class-endowed societies, when in fact it is a unique characteristic of one type of society." The dominance of production over politics defines civil society and guarantees liberty. That seems to be Gellner's bottom line. It is, as Gellner says repeatedly of Marxism, only compelling to people who wish to believe.

The Costs of Privacy: Surveillance and Reputation in America.
By Steven L. Nock. Aldine de Gruyter, 1993. 149 pp. Cloth, $37.95; paper, $17.95.

Reviewer: KAREN A. CERULO, *Rutgers University*

I remember quite vividly, as a graduate student, spending significant portions of my days reading books like Chandler's *The Visible Hand* or Beniger's *The Control Revolution*. These were socio-historical accounts that focused us on those wondrous years at the end of the 1800s — a time when shores were joined by the cold metal of train tracks and worlds were merged by the vibrating signals of telegraph wires. The final years of the nineteenth century were all about possibilities, and when I delved into pages describing those years, I remember wondering if there could have been a more exciting time to be alive.

Recent technological advancements — fiber optics, satellite transmission, cellular radio technology, etc. — have silenced the query of my graduate school days. That "more exciting time" is here. In the world of cyberspace, possibilities abound. Virtual social realities afford the potential to eradicate distance, reforge social bonds, recast social roles, and more. Yet, along with these possibilities follow the inevitable prices. Steven Nock's book, *The Costs of Privacy*, addresses one of these problems, namely, the reconstitution of social trust.

Nock's argument is based on a triangular foundation that considers trust relative to privacy and "strangerhood." He defines the modern society as one increasingly characterized by anonymous actors who cherish their privacy. Under such conditions, Nock then explores the ways in which trust, an element so vital to the maintenance of social ties, is maintained. Among "strangers," trust cannot be maintained via traditional avenues of intimacy and knowledge of other. The answer,

for Nock, lies in the more formal mechanism of surveillance; surveillance allows for an informational basis of trust.

Nock initiates his thesis by charting the growth of surveillance relative to the growth of strangerhood and privacy. He pays special attention to the growth of one specific group of strangers so prevalent in current day society — those living in independent, non-family arrangements. Using this social group as a guide, Nock takes readers through two specific means of monitoring the stranger: the establishment of credentials, and ritualistic ordeals or tests of truth. The core of the book is devoted to "case studies" of credentials and ordeals: credit records, licensing, lie detectors, drug tests, and the like. His documentation regarding the rise and use of these techniques brings issues of social control, individual liberty, and exploitation to the forefront of the reader's attention.

The book's only flaw rests in the fact that it fails to pursue many of the most interesting issues that it raises. While the book documents the surveillance phenomenon, it offers little with regard to the implications of surveillance for various aspects of social interaction and social grouping. So, for example, while Nock notes that increased surveillance can be associated with decreased risk, he fails to speculate on the consequences of such a relationship for the groups it comes to characterize. Similarly, he notes that surveillance information will come to replace one's family of orientation in the construction of personal identity. Yet again, he is silent on the many interesting ramifications of such a shift — a shift that places empirical observation above subjective interpretation of self.

Despite these minor flaws, *The Costs of Privacy* is provocative, clearly written, and builds on timely and interesting events. It holds promise for researchers working in areas such as communication, community, culture, and technology. The book represents a useful addition to graduate seminars in these areas as well. Finally, it is written in an accessible manner, and could be added to advanced undergraduate seminar courses. All in all, social scientists should find this book worth reading; in addition, it could prove a catalyst to many stimulating discussions and research initiatives.

Old Nations, New World: Conceptions of World Order.
Edited by David Jacobson. Westview Press, 1994. 235 pp. Cloth, $54.95.

Reviewer: ALVIN Y. SO, *University of Hawaii*

In the introduction, David Jacobson points to two principal developments which contributed to the emerging world order by the turn of the twentieth century. First, there is the end of the Cold War, which has brought attention to national and regional differences, and the different conceptions and idioms of world order. Second, there is the process of globalization of issues like the economy, migrants, human rights codes, and ecological deterioration. As a result of this upsurge of transnational concerns, a wealth of domestic issues have been internationalized, and international and regional institutions have concomitantly become more salient.

The chapters of the book, then, analyze how the key states have changed their understanding of the emerging world order in the 1990s. Gilbert Rozman observes that Japan is groping for a more independent role with the end of the Cold War. Furthermore, Japan perceives that economic development, rather than military prowess, is the key to world prominence.

Samuel Kim points out the irony that China invokes Westphalian principles of national sovereignty in the post-Cold War world. Given communist China's

historical conflict with the West, it is surprising that she should draw on western ideas to protect her independence.

Stephen Chiu remarks that the dramatic economic growth of Hong Kong, Singapore, South Korea, and Taiwan was encouraged by the politics of the Cold War and the liberal international economic order. Subsequently, the post-Cold War world led them to pursue new strategies of industrial restructuring and outward relocation.

Under the conditions of the Cold War, Shlomo Aronson writes, West Germany sought to rehabilitate itself through economic growth. But the end of the Cold War has brought a complex choice. Will a united Germany remain obligated to its European and Western partnership?

George Thomas notes that in the Cold War era, the U.S. defined its moral obligation as spreading freedom and democracy as well as containing communism. However, the end of the Cold War resulted in the collapse of the containment strategy and made the American identity as a moral leader problematic.

Elizabeth Wishnick writes that more than a dozen new states have risen in the ashes of the Soviet Union, all trying to establish their own national identities and to define relations with each other and the rest of the world. Each of these states confronts questions about ethnicity, religious heritage, and models of economic development.

In the concluding chapter, David Jacobson contends that the system of sovereign states is in gradual decay. The end of the Cold War and interrelated structural changes have brought regional and international institutions to the fore. In particular, Jacobson stresses that international human rights legal codes are replacing national sovereignty as the basis for state legitimacy, and NGOs (Non-Government Organizations) have become the dynamic agents that propagate, reproduce, and interpret the premises of the international order itself.

This is a timely and important book. It is indispensable for any researcher who wants to find out the latest political and cultural development in key states. The chapters are all well written and of very high quality. Nevertheless, readers may wonder whether the world order in the 1990s is really new. Global economy, international migration, human rights issues, and environmental problems have certainly existed for a long time. Even the end of the Cold War may not be that novel, as it is just another cyclical shift from a "bipolar" pattern to a "multipolar" pattern in the world-economy.

So it seems premature to assert that the present system of sovereign states is in decay. First, even in the European Community, Shlomo Aronson points out that "the nation states are still present as economic framework that exist next to, and sometimes — in terms of power — above the international level." Second, while international human rights issues are getting more important, they surely are not replacing sovereignty and national self-determination, as seen by the fact that the U.S. quickly dropped human rights issues in return for more economic ties with China. Finally, financially dependent on key states, many NGOs have been de-radicalized and their orientation has shifted from "system-transforming" to "system-maintaining" (Samuel Kim's terms). Therefore, it is doubtful that the NGOs will become the dynamic agents to promote the new world order.

Governing Capitalist Economies: Performance and Control of Economic Sectors. *Edited by J. Rogers Hollingsworth, Phillippe Schmitter, and Wolfgang Streeck.* Oxford University Press, 1994. 316 pp. Cloth, $45.00.

Reviewer: GREGORY HOOKS, *Washington State University*

The contributors to *Governing Capitalist Economies* bring a variant of institutionalism to the analysis of economic governance. Each chapter is based upon a historical comparison of two or more nations. The contributors concentrated on one sector and did not attempt a sweeping comparison of the overall national pattern of economic governance. The case studies include: circuit boards in Japan and the U.S. (O'Brien); shipbuilding in Germany, Sweden, and Japan (Strath); machine tools in Germany and the U.S. (Herrigel); chemicals in several industrial democracies (Grant and Paterson); automobiles in Germany, Great Britain and France (Dankbaar); dairy in Austria, Germany, and Great Britain (Traxler and Unger); consumer electronics in Britain and France (Cawson); and securities in Canada, Great Britain, and the U.S. (Coleman). The editors' beginning and concluding chapters develop a (loose) conceptual framework and provided an overview of the volume. Despite the unavoidable centrifugal tendencies in edited volumes, the editors' conceptual framework is reflected in each of the chapters — to a greater or lesser degree.

The contributors build on recent efforts to synthesize economics and sociology. "The common premise throughout this book is that economic action is a special case of social action, and therefore, needs to be governed by institutional arrangements." Governance refers to a system of institutional arrangements — including rules and the enforcement of these rules — that regulates economic activities. The contributors are not only interested in identifying the institutions and the processes that serve to reproduce or transform them, they are also concerned with when and why institutional arrangements "work." In this effort, the contributors identify the stated goals of economic governance and examine the trends in market shares, profits, and innovation in the cases under investigation. The conception of institutions is broadened to encompass a variety of institutional forms that have appeared in the social sciences literature: markets and corporate hierarchies, informal networks, states, and associations (including private interest government). The collaborators do not put forward a grand theory that defines the relative importance and the relationship among these various institutional alternatives. Rather, they emphasize the variation in the presence, importance, and relationship among these institutions across nations and over time. The comparative case studies demonstrate that forms of economic governance are not determined by technologies, nor are they dictated by profit maximization strategies in any simple fashion. Rather, institutional arrangements emerge from the nation's political and economic history, and institutional inertia channels economic activity.

Those looking for a tightly woven theoretical framework applied to diverse cases will be dissatisfied with this volume. However, those seeking a focused but theoretically eclectic examination of historical cases of economic governance will find an interesting resource. In the tradition of Weberian "thick description," the case studies raise a number of puzzling theoretical issues, but offer few final answers. In fact, the most interesting insights were not driven by the theory, but instead were fortuitous aspects of relating the historical developments of the cases under examination. Although the editors discuss "convergence" theory — from its early postwar emphasis on the merging of capitalism and socialism to its current emphasis on globalization and declining sovereignty of states — that discussion does not offer a new theoretical synthesis. However, Coleman's examination of the securities industry in Canada, the United Kingdom, and the U.S. makes this issue

salient and interesting. His case study documents the manner in which the globalization of the securities market has overrun the distinctive governance regimes of these three Anglo-Saxon capitalist democracies — and suggests that the distinctive German and Japanese financial systems may soon confront a crisis of governance in this increasingly internationalized sector.

I recommend the book to those contemplating case studies of economic policies in the advanced capitalist nations. Those who would put forward a grand theoretical synthesis of economic governance would do well to read this book. One of the most important contributions of this book is its forceful reminder that history is messy and inefficient.

Forgotten Places: Uneven Development in Rural America.
Edited by Thomas A. Lyson and William W. Falk. University Press of Kansas, 1993. 278 pp. Cloth, $39.00; paper, $14.95.

Reviewer: RONALD C. WIMBERLEY, *North Carolina State University*

Flying across the U.S. at 35,000 feet is not a normal place to draft a book review, but it does provide a unique vantage for thinking about forgotten places and looking for them below. A couple of observations emerge. First, forgotten places are not just out of mind, they are mostly out of sight as well. Second, when they are spotted, they are all but lost in the vast rural space of our country's geography.

Lyson and Falk have not only found some forgotten places, they have found several rural social scientists to give us a close-up look. *Forgotten Places* is a contribution of the Rural Sociological Society's Task Force on Persistent Rural Poverty. It contains nine case studies on rural poverty with an introduction and conclusion by the editors.

The *forgotten* places brought into view by this book are actually *poor* forgotten places. And beyond the plane of forgetfulness and poverty, a third focus — rurality — rises to make the subject three dimensional and to show more clearly the "uneven development" of the socioeconomic terrain. Altogether, the forgotten places called to our attention are poor forgotten *rural* places.

Few of the forgotten places can be found in the central area of the U.S. Other than Appalachia, the Mississippi Delta, and the Missouri Ozarks, the rest are at the geographical as well as rural, social, and economic peripheries: in northern New England, Michigan's upper peninsula, the Black Belt South, the Rio Grande Valley, rural California, and the Oregon timber country. An example mentioned but unused is also near the border: the Indian reservations of the Dakotas.

My campus colleague Jim Clark asks why so much of North Carolina's rural poverty falls so near the state lines. Why so many poor forgotten rural places fall at the nation's borders is another good question, but one for another book. This book does enough already.

Lyson and Falk's introduction and conclusion provide, as usual, interesting reading. In what I think is an especially important chapter, Falk, Clarence Talley, and Bruce Rankin make the point that while rural areas tend to be poorer than urban areas, there is also inequality within regions between the Black Belt and the South, for example. "Those living in the Black Belt are more likely to have less education and lower incomes and to suffer from poverty and unemployment," and the economic gap appears to be widening. They conclude, "No part of the United States has been more forgotten or overlooked than the Black Belt." The inequalities of both rural residence and region, the authors note, make such places as the Black Belt "doubly disadvantaged."

Besides being the thesis of the Black Belt chapter, the multiple disadvantages of poor forgotten rural places become the major theme of the book. Other chapters describe how such disadvantages as poverty and rurality appear in other regions. In a rich and lucid chapter on California's rural communities, for example, Ted Bradshaw shows further disadvantages within rural places themselves. He sees a process of rural bifurcation where new, upscale, urban-based growth invades a rural area and forces the old, local, indigenous, rural population into worse conditions. The new development attracts attention; the old rural culture becomes less visible and more forgotten.

The book reveals more about America's poor forgotten rural places than one can observe from the vantage of a cross-country flight and costs quite a bit less. But whether you fly or not, don't forget the book. It's the rural binoculars that help to find and see the forgotten places.

Appalachia in an International Context: Cross-National Comparisons of Developing Regions.
Edited by Phillip J. Obermiller and William W. Philliber. Praeger, 1994. 238 pp. Cloth, $55.00.

Reviewer: ED KNIPE, *Virginia Commonwealth University*

This edited volume brings together several previously published works which compare Appalachia with similar communities and regions elsewhere.

In the introduction, William W. Philliber critically reviews several different approaches to defining the Appalachian region in particular and their implications for studying regionalism in general. Without so stating, the rest of this volume reflects those approaches within an international context. For example, the psychological basis of regionalism is illustrated by John Stephenson's comparison of "Shiloh," a community in North Carolina with Ford, a small community in Scotland. In both of these communities migrations are attributed to a search for what he calls "place," a search that reflects dissatisfaction with one's present residence. Susan Abbott compares school children in eastern Kentucky and Kenya on anxiety symptoms and gender expectations, which she concludes are similar because of similar community structures. Nelda K. Pearson compares the characteristics of community-based problem-solving groups in southwest Virginia and Newfoundland. She finds that these groups function differently as a consequence of the motives of "prime initiators" (those who founded the groups) and differences in Canadian and U.S. cultures.

Several chapters exemplify the economic approach to region. Graham Day compares Wales and central Appalachia as following parallel but somewhat different post-World War II histories of economic development. Richard A. Couto analyzes the antecedent conditions leading to the 1977-1978 coal miners' strike in Appalachia and the British miners' strike of 1984-1985. He suggests that both are the result of postindustrial conditions in these industries as measured by the rise of intermediate organizations that attempt to compensate for lost social overhead. The strike in both instances is a dramatic expression of that loss. John Gaventa traces the movement of a single industry — a seat belt manufacturer — from Michigan to Knoxville, Tennessee, then to Alabama and finally to Mexico to assess the negative impact on workers in Knoxville. Benita J. Howell argues that the stereotypical view of Appalachians as "backward" is not an isolated phenomenon. Such views of various populations may indeed be universal and may have their roots in economic ties between a region and the larger social and cultural matrix within which it is found.

Although not directly related to analytic disputes concerning definitions of a region, several chapters focus on attempts to ameliorate "problems" associated with Appalachia. For example, Richard A. Couto paints a rather dismal picture of the future for a deindustrialized and postindustrialized America based on what has happened and is happening in Appalachia. His argument rests on defining the region in terms of labor and the kinds of work that mark a shift from an industrial economy. Glenn A. Mitchell compares Catalonia and Appalachia on the basis of what he calls "minority regions" and finds that differences in development are the consequence of historical variations in political and cultural characteristics of each region. Phillip J. Obermiller compares regional planning organizations in southern Italy and Appalachia. Assuming Appalachia is a region, he argues that both areas have failed to improve the conditions for which they were organized and suggests that grassroots efforts are the only way these regions can be developed. And finally, Peter R. Sinclair reviews several theoretical models that have been used to understand regional inequity in Canada. He specifically focuses on the maritimes as his most inequitable regions, but fails to mention Appalachia.

While each chapter offers interesting insights into Appalachia, the book needs a summary chapter to find those common threads that tie these disparate studies together. The failure is annoyingly common in books of readings. I think of a book as a tool, the usefulness of which is reflected in a good index and bibliography. While this index is adequate, many of the references cited in the text could not be found in the bibliography. For serious scholars this omission can be maddening.

Relations into Rhetorics: Local Elite Social Structure in Norfolk, England, 1540-1640.
By Peter S. Bearman. Rutgers University Press, 1993. 216 pp. $40.00.

Reviewer: RICHARD LACHMANN, *State University of New York at Albany*

This is a book which you might want to give to your historian friends since it presents sociology in the best possible light. Bearman addresses the most intractable and the most written about issue in English historiography: the question of why country-based gentry sided with Parliament or with King Charles I in the English Revolution and Civil War of 1640-1649.

This question has driven many contemporary scholars to the point of despair. A group of prominent revisionists now declares that, because people had various reasons to choose sides and even more reasons to ally with both sides or try to remain neutral in the Civil War, that conflict is beyond understanding. Bearman cuts through such antirationalist defeatism. Bearman, like the revisionists, considers the simple Marxist class analysis of the Civil war "clearly wrong." One response to that failure has been to draw ever finer distinctions among class fractions in hopes of finding specific "categorical affiliations" which could predict action. Another direction, favored by the revisionists, has been to examine the individual biographies of participants in the Civil War in hopes of identifying idiosyncratic rather than categorical bases for actions.

Bearman proposes another, specifically sociological, strategy. He sees "structural equivalence of individuals in social networks," rather than common attributes, as the bases for interests and actions. He uses blockmodels to go beyond the revisionists' concern with individuals' particular social relations and to aggregate "individuals with similar patterns of social relationships into an equivalency class." Network analyses allow him to test the saliency of different social factors, identified by Marxists and others, as the bases for action in each historical era.

The blockmodel analyses reveal that kinship in the sixteenth century and religious patronage in the decades leading up to the Civil War were far more important than the usual factors emphasized in historical research on the Civil War — land holdings, religious identity, officeholding, connections to the royal court, and residence in counties — in determining social and political identities among the ruling elites of Norfolk. (Bearman's analysis is confined to the county of Norfolk; however, he is exemplary in identifying Norfolk's peculiarities and in explaining the extent to which his findings can be generalized for the rest of England.)

Kinship determined political affiliation until the late sixteenth century, when it declined as the principal organizing relationship in Norfolk. Kinship lost importance when and where the crown succeeded in breaking the key magnate dynasties around which kin blocks were organized. At the same time gentry excluded from magnate blocks sought status and power by marrying outside of Norfolk, further reducing the cohesive and stratifying force of kinship.

State formation was a result of the efforts of gentry from below to gain status as well as the strategies of monarchs from above. The gentry sought royal favor both for financial advantage and to replace magnate-centered kin networks as an organizing basis of county politics. The crown's generous granting of favors shattered magnate authority but at the cost of surrendering control of future patronage decisions to factions organized from the counties. Such factions did not provide clear bases for stratification within counties. Bearman describes institutional paralysis on the Norfolk Commission of the Peace (the county's main governing body) during the first decades of the seventeenth century.

Norfolk gentry overcame status and political confusion through the use of religious patronage. Gentry came to hold a majority of advowsons (the right to appoint parish ministers) in the aftermath of the Henrician Reformation. Spurred by Puritan zeal, a minority of gentry started using advowsons to control the ideology of ministers rather than just to profit from parish revenues as had most gentry in the late sixteenth century. Orthodox and Catholic patrons responded with ideological appointments of their own. Networks of gentry linked through the appointment of the same ministers (either simultaneously or in sequence) reveal blocks defined increasingly by religious ideology in the decades leading up to the Civil War. The blocks of religious patronage are the best predictors (better than an individual's own religious identity, ties to the court, or occupation) of affiliation and activism during the Civil War.

Bearman uses sociological techniques and a structural sensibility to produce a model which is far more predictive of the loyalties of Norfolk gentry in the Civil War than Marxist and revisionist historians who are unable to settle upon a method to join theoretical assertions with the muck of detail. Applied to all of England, Bearman's approach would go far in settling the seemingly intractable debates over the Civil War.

Bearman's book should be a tonic to sociologists as well. He reminds us of the value of our discipline. Our understanding of social structure allows us to "abstract from context to yield pattern." Those patterns in turn allow us to model identity and thereby impute "motive to actors on the basis of the position that they occupy in a network of positions." Bearman's methodological precision and theoretical clarity provide a model for sociologists and for historians.

Art Crime.
By John E. Conklin. Praeger, 1994. 322 pp. Cloth, $49.95.

Reviewer: KURT LANG, *University of Washington*

Almost overwhelming in its array of synopses of specific "art crimes," with biographies of perpetrators and victims, descriptions of techniques typical of art thieves and art forgers and of situations that tempt them, this book by John E. Conklin, professor of sociology at Tufts University and author of several previous criminology books, can be recommended for the fast-moving kaleidoscope it offers of just about everything you ever wanted to know about this diminishingly arcane subject. Yet, intriguing as these vignettes may be to the novice, anyone with some familiarity with the trade in art and antiques is unlikely to learn much from this book beyond some titillating details.

Data come mostly from news reports and from a handful of journals for collectors. Other material, particularly about the plunder and trade in antiquities, comes from *The Art Crisis* (1975), a mildly popular book by Bonnie Burnham. But Conklin's acknowledged reliance on such sources is not, for this reviewer, the major shortcoming of what, despite its well-paced movement from topic to topic, strikes me as scissors-and-paste sociology.

Not that *Art Crime* comes without the requisite "theoretical" introduction. It reviews some high spots in the literature on the social construction of value in art (also the subject of a separate chapter), on the motives underlying criminal behavior, on the neutralization stratagems by which such behavior is justified, and on the social organization of art crime. One only wishes that these subjects had been pursued more diligently and more systematically throughout a book that ends with a menu for curbing crime in this often speculative and potentially lucrative market.

It stands to reason, therefore, that many of the crimes most typical of the art trade — insurance fraud, bad checks, worthless collateral, violation of export restrictions, failure to acknowledge conflicts of interest — are also found in other areas of business. To be sure, Conklin is right on target in noting certain peculiarities of the art trade, among them tenuous nature of artistic value, the trust assumed in transactions, and the inexpertness of most purchasers. Nor is there any real parallel in the business world to art vandalism, to which he devotes a chapter. Nevertheless, in the opinion of this reviewer, who makes no claim to special expertise in criminology, the sociological relevance of this lively book could have been improved by some explicit references to the generic categories of fraud and theft into which these art crimes can fruitfully be placed.

More illuminating is the discussion of the mentalities of the perpetrators of art crimes. What I find lacking, on the other hand, is correspondingly detailed treatment of the legitimate artist as either an offender or victim. Creative individuals are often pushed into "criminality" by the censorship policies of repressive regimes. Faced with the export ban imposed on all but a few favored artists in the former Soviet bloc, many of them managed to smuggle their work out of the country. Almost every artist I interviewed last fall in what used to be the German Democratic Republic proudly told of some illegality by which he or she had outwitted the authorities. And, finally, while *Art Crime* tells us a good deal about the victimization of artists by gallery owners, it barely touches on the many infringements of artists' copyrights, about which they are forced forever to be on the defensive. Unfortunately such ordinary crimes rarely make journalistic headlines. This book, too, in its selection and treatment of examples seems to have been driven more by their newsworthiness than by their sociological relevance.

Inequality, Crime, and Social Control.
Edited by George S. Bridges and Martha A. Myers. Westview, 1994. 330 pp. Cloth, $59.95; paper, $33.95.

Reviewer: BARBARA J. COSTELLO, *Mississippi State University*

This work is a collection of sixteen essays developed from a 1992 University of Georgia conference. The essays are organized around four themes: (1) the extension or integration of social control theories, (2) the refinement of conceptualizations of inequality and its relationship to crime and social control, (3) the specification of the nature and direction of relationships among various forms of social control, and (4) analysis of the processes by which individual and group values come to be reflected in law and official reactions to deviance.

The major strength of this work lies in the diversity of new theoretical approaches it presents. Tittle's essay provides a compelling argument for theoretical integration by pointing out that four seemingly divergent perspectives (consensus, conflict, social psychological, and bureaucratic) all predict inequality in sanctioning, though each contains flaws precluding the complete explanation of this inequality. Liska argues for the refinement of conflict theory, which as currently formulated and tested "explains everything and predicts nothing." He bases this claim, in part, on the fact that conflict theory can explain apparently contradictory empirical findings. For example, increasing organization of minority groups can lead to an increasing threat to the majority and consequent increases in social control, but it can also lead to decreases in social control through the minority's increased ability to resist control. Liska also criticizes the disjointed nature of the literature on conflict theory, which is organized by form of social control rather than by theoretical propositions. Platt's analysis of recent developments in the field echoes many of Liska's concerns and calls for abandoning the case study approach in favor of a broader, multidisciplinary approach to social control issues. The final contribution in the general theory section is Peter Manning's innovative application of Bourdieu's structuralist theory to the policing of domestic violence.

This diversity continues throughout the remainder of the book, which contains additional theoretical developments as well as empirical analyses. Unfortunately, the diversity that is the strength of the work is also the book's major weakness, as the authors often ignore each other's conclusions and recommendations. For example, Kathleen Daly argues that studies of gender differences in sentencing, including her own prior work, have overlooked subtle differences between the seriousness of men's and women's offenses that can explain virtually all observed sentencing "disparity." In a subsequent chapter, Bridges and Beretta cite Daly's prior work as evidence that gender differences in crime cannot explain gender differences in imprisonment. They further argue that "explanations of disparities in punishment must move away from perspectives that rely heavily on differences between men and women in criminal behavior."

Tolnay and Beck's analysis of Southern black lynchings does not ignore earlier recommendations but seems to misunderstand them. They cite Tittle's argument for the integration of consensus and conflict perspectives, implying that his recommendations are valid, yet go on to state that "there can be little doubt that Southern whites shared a strong consensus that blacks were . . . a threat to the integrity of their cherished society." They therefore misinterpret consensus theory as implying that consensus exists only *within* conflicting interest groups, which is actually an assumption made by conflict theorists.

These contradictions would not be so troubling if they were presented as points of contention, but Bridges and Beretta do not argue against Daly's current perspec-

tive and Tolnay and Beck do not explicitly criticize Tittle's arguments. Granted, it is difficult to compile sixteen papers into a completely coherent volume, especially when the goal is to present innovations in a field. Nonetheless, I was left wondering who will be likely to follow the authors' recommendations if not their fellow contributors.

Crime in the Making: Pathways and Turning Points Through Life.
By Robert J. Sampson and John H. Laub. Harvard University Press, 1993. 309 pp. $32.50.

Reviewer: ROLAND CHILTON, *University of Massachusetts at Amherst*

This book reexamines delinquency and crime data collected by Sheldon and Eleanor Glueck from 1939 to 1961 and is a benchmark for current and future longitudinal studies of crime and delinquency. After recasting the data, they were able to reassess much of the criticism of the Gluecks' book, *Unraveling Juvenile Delinquency* (1950). More importantly, they use the data to illuminate and advance several current theoretical and empirical issues in criminology.

In the opening chapters, Sampson and Laub present (1) a theoretical framework, (2) the Gluecks' perspective, (3) a detailed description of the Gluecks' study, and (4) a description of the procedures used to reorganize the basic data. Their discussion of the work and the backgrounds of the Gluecks provides a convincing explanation for the limited impact of *Unraveling Juvenile Delinquency* on main stream criminology. Subsequent chapters reassess the importance of informal family control and school, peer, and sibling influences by looking at the probable contribution of structural factors, process variables, and propensity measures to official and unofficial delinquency. Their basic conclusion after these initial analyses is that "when the bonds linking a youth to a society — whether through the family or school — are weakened, the probability of delinquency is increased."

However, perhaps the most important results are found in the chapters focused on the continuity of behavior over time and the impact of social bonds on adult behavior. They find continuity in official criminal behavior over a thirty-year period but then explore the possibility that adult social bonds explain variations in adult crime beyond those explained by childhood differences. After exploring this possibility, they conclude that "childhood delinquency and adult social bonds (job stability and marital attachment) independently explain significant variations in adult crime."

The book has two additional components, a qualitative analysis of some of the voluminous files created by the Gluecks' research staff and transcripts of interviews with three of these staff members. There were several advantages to using data collected by the Gluecks, including the detail and the range of information sources, the availability of self-reported, parent-reported, and teacher-reported delinquency, and a very successful follow-up effort. However, a major limitation of *Crime in the Making* is the limitation produced by the Gluecks' original research design. It is a comparison of 500 young white males held in two Massachusetts prisons for juveniles with 500 nonincarcerated young white males who were living in poor communities in Boston in 1939. Most lived through a great depression and were followed through a great war and a postwar period. Sampson and Laub argue that this is a strength of the study because it provides an important comparative base for assessing current concerns of race, crime, and the underclass.

I think the data would have been more useful if it had *not* been limited to white males in poor neighborhoods. It would have been useful then, and would be today,

to know why so little crime is committed by girls and women. Moreover, the Gluecks truncated the social class distribution when they focused on poor neighborhoods and excluded black delinquents. As a result any conclusions about the impact of social class are weakened.

This limitation aside, the book is an important effort that will be of great value to criminologists and criminology students for years to come. This is primarily because of the authors' willingness to examine the impact of structural, interactional, and propensity factors in the production of delinquency and crime. *Crime in the Making* is an important and well-crafted explanation of why some children become delinquent, why some adults become criminal, and why only some of these children and adults continue to engage in criminal conduct as they mature.

Wannabe: Gangs in Suburbs and Schools.
By Daniel J. Monti. Blackwell, 1994. 174 pp. Cloth, $16.95.

Reviewer: MALCOLM KLEIN, *University of Southern California*

With the proliferation of gangs of many sorts across the country, there has emerged a matching proliferation of new research studies of gangs. With a few notable exceptions, these new studies are case studies either of a single gang or of a single city with gangs. Generalizable knowledge thus depends on the ability and willingness of case study authors to describe the parameters of these gangs or gang situations, and to relate these to what has been revealed in prior published work. *Wannabe,* a report on suburban gangs in a midwestern metropolitan area based on youth interviews in elementary, junior, and senior high schools, is deficient on both criteria for deriving generalizable knowledge. Although engagingly written and sprinkled with interesting quotes taken from the interviews, the work is of indeterminate value.

While he acknowledges the ambiguities inherent in defining or depicting gangs, Monti offers little by way of definitional clarity. Because his interviews are with unknown proportions of nongang and gang respondents, ranging from elementary to high school ages, the boundaries of "gangness" are hard to discern. The book's title belies the seriousness of involvement of a number of these respondents and other gang members they describe.

Research is limited to interviews in school, often with a counselor present. No corroboration of interview reports is attempted via official data, nor by street or other observation, nor by interviews with teachers, other officials, parents, or dropouts. Also, respondents were warned not to speak of their delinquent offenses (though some did). Thus we are offered no crime data, no demographics, no analysis of differences between gang members and nonmembers, little on gang histories or structures, etc. The problems are exacerbated by what, to this reviewer, seem to be rather intrusive and leading interview procedures.

Until the last chapter on policy issues, which cites but makes minimal use of prior research, there is virtually no scholarship in this work. Presumably this can in part be attributed to the attempt to reach a wide lay audience, but it also reflects Monti's view on prior research as "social scientific claptrap" and "drivel." In the final chapter he characterizes his fellow social scientists as "a sociologist, psychologist, or any other slick expert pushing an open hand and an empty head."

This disdain for and ignoring of prior works leads Monti to some serious errors. He claims to be the first to study gangs in suburbs; he says that independent female gangs are a new phenomenon; he seems surprised to find gang members coming from middle-class, two-parent families; he concludes that "gangs are democratic to

a fault," being equally open to all ethnicities, and equally open to middle- and lower-class youth; he concludes that where gangs are formed is unimportant, characterizing other work on this issue as existing only "in Frederic Thrasher's mind. All the researchers who faithfully followed the path he cut made the same mistake and were as lost as he was."

Failure to follow the literature on the relationship between street gangs and drug sales leads Monti to a series of confused and contradictory descriptions of this relationship, and of the place of migratory Bloods and Crips in spreading drug markets. Indeed, when his respondents offer contrasting views on drug issues and on the importance of big-city gangs in his suburbs, he is at a loss even to attempt any resolutions. Other scholars' reports and other methods could have helped him in such circumstances.

In sum, another opportunity to add to our general knowledge of gangs has been missed. If Monti had presented his materials in the spirit of exploration, this review could have been more positive. Instead, he presents his findings and conclusions as rather definitive knowledge about a heretofore undiscovered phenomenon. Since we are not provided with replicable methods or data, such knowledge would seem to be on shaky ground. This reviewer is not satisfied with "revealed truths."

Street Kids, Street Drugs, Street Crime: An Examination of Drug Use and Serious Delinquency in Miami.
By James A. Inciardi, Ruth Horowitz, and Anne E. Pottieger. Wadsworth, 1993. 234 pp.

Reviewer: HILARY SURRATT, *University of Miami School of Medicine*

This study provides a much needed account of the drug use and criminal involvement of a large cohort of seriously delinquent youths in the Miami, Florida, area. The authors' use of street research techniques provides empirical data on a hard-to-reach population that has been understudied by traditional delinquency research. The beginning of the book presents a comprehensive overview of traditional delinquency theories including social disorganization, control theory, and social learning theory, but asserts that none of these perspectives is sufficient to explain serious juvenile involvement in crime and drugs. While much media attention has been focused on young crack users since the mid-1980s, little empirical research has been conducted on adolescent criminality and drug use and therein lies the contribution of *Street Kids, Street Drugs, Street Crime.* The authors provide in-depth drug use and criminal histories for a multi-ethnic sample of 611 serious delinquents who exhibit a pattern of drug use from alcohol, to marijuana, to powder cocaine, and finally to crack-cocaine. The vast majority of these were found to be polydrug users who began using crack at the alarmingly early age of 14 years. Despite the magnitude of the adolescent drug problem, the authors point out that there remains a serious lack of drug treatment facilities for juveniles. Only 4% of this sample had ever been in a drug treatment program, which presents a strong argument for compulsory drug treatment of those coming to the attention of the criminal justice system.

These delinquents reported an astonishing number of criminal offenses in the previous twelve months. Drug business offenses, especially retail drug sales, were reported by almost every respondent and a staggering 88% of the female respondents had engaged in some form of prostitution. Importantly, while previous research has suggested that black males commit a disproportionate amount of serious juvenile crime, the findings of this study do not support such a conclusion. The authors found that the *overall* crime level does not vary significantly by either

gender or ethnicity, and they emphasize that crack use and crime are not just problems among inner-city, minority juveniles but among white working and middle-class young people as well.

In the closing chapter of the book, the authors present what is perhaps the most important public health implication of this research: the potential for HIV transmission. Although reported cases of AIDS among adolescents are infrequent, the introduction of crack-cocaine among adolescent drug users may be a contributing factor to infection with HIV. The use of crack has been associated with higher levels of sexual activity, through prostitution and barter of sex for crack. In addition, a significant number of the 611 young people interviewed in this study were engaging in behaviors that put them at risk for HIV infection.

Overall, this is an important book for both students and researchers in the areas of delinquency, criminology, and substance abuse, because of the contribution it makes to our knowledge of chronic criminal behavior and drug involvement among American youth.

Childerley: Nature and Morality in a Country Village.
By Michael Mayerfeld Bell. University of Chicago Press, 1994. xi + 279 pp. Cloth, $32.50.

Reviewer: GEORGE A. HILLERY JR., *Virginia Polytechnic Institute and State University*

This an excellent book. Well written and well organized, it is a pleasure to read. To a large extent, Bell lets the people tell their stories about things of interest to *them*, and this helps make them interesting to us.

Childerley is an exurban English village of 475 inhabitants, about two hours from London. Since it is composed heavily of those who have rejected the city (at least as a domicile) and the anonymity of mass society, we are surprised to learn that "few Childerleyans know more than a small fraction of the other residents in the village." Yet the village is a relatively coherent entity.

Bell explores three main aspects of this entity: social class, the attitudes of the villagers toward nature, and the interrelatedness of village traits. Class is given most attention. Though there are numerous status variations in wealth, traditional social class (nobility and commoner), and occupation, the villagers usually think in terms of two levels. When visiting, who uses the front door and who the back door? The importance of class, however, is declining, as suggested by the people's distaste for it.

Bell deals with ecology in its original meaning: the relation of organisms to their environment. This approach shows especially in his treatment of the geography of the countryside and the relation of the people to it. Today there may be in fact little difference between country and city, but regardless of the facts, "every village resident agrees that there is such a thing as the country, distinct from the city — and every village resident agrees that this distinction is important." Many have come to Childerley precisely because of their desire for country life.

An impressive feature of the book is the emphasis on interrelatedness of village traits. Bell takes several key phenomena and shows how they influence what goes on in Childerley. Class especially is tied in with most of village life. We are given, for example, a clearer picture of lower and upper class relationships by a description of class-based differences in attitudes toward the fox hunt.

Bell achieves something not often found in sociology — a treatment of society both from the point of view of individuals and from the perspective of the society

and its institutions. He does this by selecting for intensive interviews representatives from each class and gender. The result does much to bridge the micro-macro gap.

More explicit emphasis is placed on gender than on family, which is interesting since the family has been a mainstay of class. Although families are mentioned frequently, Bell apparently feels that family is not as important as it once was. To understand community behavior, gender may indeed become more significant as family and group interests give way to individualism. Nevertheless, according to the descriptions Bell gives, family still explains much important behavior and could have been used more explicitly as an explanatory variable.

The strength of a book is often found in its weaknesses. *Childerley* has few. Though not a structural-functionalist, Bell nevertheless assumes that everything is related to everything else. Indeed, he gives this idea fresh perspective by introducing the concept of "resonance." Taking as his model the change ringing of bells, he suggests that ideas resonate with each other just as bells do. The model will probably prove useful as long as one does not follow it too closely. Human groups are not put together like bells. Their bonding is primarily symbolic, which means that connections between people are essentially arbitrary and basically indeterminant.

But when considered from a broader perspective, criticisms offered here will not appreciably detract from this truly outstanding work.

The Quest for the Other: Ethnic Tourism in San Cristobal, Mexico.
By Pierre L. van den Berghe. University of Washington Press, 1994. xii + 169 pp. $24.95.

Reviewer: ERIK COHEN, *The Hebrew University of Jerusalem*

San Cristobal is the center of ethnic tourism in highland Chiapas — and according to the author, a rare example of successful, small-scale, environmentally friendly tourist development. It is a *ladino* town, surrounded by a low-class, peasant stratum of Indian micronations that are the alleged goal of the tourists' quest for the "Other." Alas, it is also at the center of the Zapatista Indian uprising, and was briefly occupied by the rebel army in early 1994, about the time this book went to press. The coincidence is somewhat ironic, particularly since the optimistic tone of the book (written by an anthropologist with decades of experience in the area, who approaches tourism from the perspective of ethnic relations theory) misses any indications of the coming storm, even though the rebellion seems to have been under preparation for as long as eight years.

The omission is hardly a failure of the author, however, who went to San Cristobal for a three-month period of fieldwork in 1990, explicitly to study tourism. But it is an indication of the depth of the gulf still separating the urban *ladino* society, among whom the author lived, from the secretive Indian communities in the countryside; it is also an indication of the difficulty any student of tourism experiences in penetrating the front which even unassuming natives present to tourists and other strangers, and behind which frustration, despair, and animosity are frequently concealed. The Indians are described as dignified, but not unfriendly; the pent-up tensions break through only in the report of one incident — the killing, some years ago, of a foreign tourist in an Indian village, after he had disregarded repeated warnings not to take photographs. The uprising did not necessarily start in the communities visited by tourists; but it would be interesting to know whether tourism played a role in triggering it off.

The book is essentially an anthropological survey rather than an in-depth community study. The author describes in detail the five principal components of tourism in San Cristobal: the tourist attractions and facilities, the mainly *ladino* inhabitants of the town, the "tourees," (i.e., the Indians), and the tourists themselves. He presents some statistical summaries of his interviews with the tourists, as well as 25 "vignettes," brief descriptions of individual visitors, in order to illustrate their diversity. The Indians, rather than the urban *ladinos*, are the goal of ethnic tourism. They are the Other. But the *ladinos* are the owners of the tourist facilities and act as middlemen between the tourists and the tourees, reaping the principal benefits of the tourist business. The Indians, at the bottom of the traditional ethnic hierarchy, benefit principally from the sale of handicrafts and from begging in towns, where most of the contacts between tourists and Indians take place; Indian communities in the countryside, except one or two which became sightseeing attractions, benefit little from tourism.

Considering that this is the first book-length study of tourism in a Mexican town, it makes informative, though not particularly exciting, reading. Neither is there any significant theoretical innovation: the author analyses his data in terms of MacCannell's, conception of "staged authenticity," but he seems somewhat ambivalent towards such staging: while he discusses it critically in the body of the text, in the conclusions he advises covert staging as a strategy to preserve the ambience of the town.

Van den Berghe finds that tourism has improved the *ladinos'* attitude toward the Indians and their appreciation of Indian culture; it appears to have had an ameliorating influence on the rigid hierarchical ethnic relations of the highlands. But the Indians' position in the hierarchy has not changed significantly; and they have remained the inscrutable Other not only for the tourist but apparently also for the author. Indeed, the concept of the Other remains opaque throughout the book. We learn little about the meaning of the Indians' "otherness" for the tourists, and the author does not attempt to penetrate it for himself; if he did, he would perhaps be able to discover the seeds of the revolt hiding behind the impassive face of the Indian "tourees."

After Ellis Island: Newcomers and Natives in the 1910 Census.
Edited by Susan Cotts Watkins. Russell Sage Foundation, 1994. xviii + 451 pp. $49.95

Reviewer: TERESA A. SULLIVAN, *The University of Texas at Austin*

The nation's census is an investment in infrastructure that is valuable not only for the immediate concerns of reapportionment, marketing, and planning, but also as an incomparable portrait of the country that can be revisited again and again. This volume shows how a census can pay dividends long after its completion. Using the Public Use Sample of the 1910 U.S. Census of Population, the demographers who wrote these chapters examine child mortality, ethnic fertility differences, household structure, residential segregation, schooling and industrial affiliation of workers at the end of the first decade of this century.

The Public Use Sample is a national random sample of the manuscript census. Larger than the comparable 1900 data base, it allows for a richer analysis of relatively small immigrant groups. Most of the data were previously unpublished. The Census asked women about the number of children they had ever borne and the number surviving, but the survival data were never before analyzed. A rather lengthy appendix provides considerable detail on the sample, and Susan Watkins's chapter provides background on the 1910 census.

The Thirteenth Census reflected the concerns then, as now, with immigration and its effects. Most of the chapters in this volume take advantage of the variables of country-of-origin and mother-tongue to examine differences between native-born and foreign-born, among generations, and among linguistic groups. Published tabulations of the era, by contrast, often reported the "foreign-born" as a group, effectively masking the differences that later chapters of this volume reveal.

One set of intriguing findings concerns the child mortality of various nationality groups. The foreign-born population had relatively high child mortality, compared with the native whites, probably due to immigrants' concentration in large cities. French-Canadian child mortality was extremely high, and Jewish child mortality was quite low, anomalous results that might arise from breast feeding practices, according to the chapter's authors, Samuel H. Preston, Douglas Ewbank, and Mark Hereward. Second-generation Americans had lower child mortality than immigrants, and wives who married in the U.S. lost fewer children to early death than women who had married abroad. Child mortality among black Americans was very high, with a mortality index that began at 0.61 higher than native whites. Even after controlling for other variables measurable in the census, the index remained elevated by 0.41, a fact that might be attributed to the very marginal economic position of blacks forty-five years after the end of the Civil War.

The contrast between black Americans and immigrants is an important part of the national portrait in 1910. Ann R. Miller, in a chapter on the industrial affiliation of immigrants, argues that the industrial transformation of the U.S. labor force was made possible by the flow of immigrant labor to the cities. Over 30% of foreign-born whites were employed in agriculture, compared with 15% of native-born whites and fewer than 7% of blacks. The continued settlement of the western states, by contrast, was disproportionately the work of native-born whites, for whom agriculture remained an important industry. Considerable specialization among the other groups was evident, perhaps because of chain migration. Central Europeans were overrepresented in mining, where over 13% of them were employed, and in manufacturing, where nearly 45% were employed. Irish and Italians were overrepresented in transportation, and Eastern Europeans were overrepresented in trade.

Not so theoretically focused as Stanley Lieberson's *A Piece of the Pie*, this volume is a signal contribution to the literature because of its rich new data base and sophisticated modeling. Given the renewed significance of immigration as this century ends, this scholarship has the added potential of shedding light on current controversies. Finally, this volume represents a cautionary tale for our current Congress. Congress did not restrict the Thirteenth Census to a seven-item short form; 85 years later, we are benefiting from the wisdom of that decision. What we will know about ourselves at the turn of the twenty-first century depends upon our willingness to make the same kind of investment in ourselves now.

The Post-War Generation and Establishment Religion: Cross-Cultural Perspectives. *Edited by Wade Clark Roof, Jackson W. Carroll, and David A. Roozen.* Westview Press, 1995. xx + 291 pp. $55.00.

Reviewer: KEVIN J. CHRISTIANO, *University of Notre Dame*

A great many words have been written in newspapers and magazines and carried over the airwaves during the last decade about an unexpected return to conventional religious practice by members of the generation born between 1946 and 1965, the so-called "baby boomers" who now dominate Western societies and in America account for one third of the total population. Having endured an adolescence of

material affluence combined with cultural unease, and after an early adulthood filled with history-making social turmoil, these scarred veterans of the sixties have quietly crossed the "generation gap," the story goes, and are ready finally for property-holding, parenthood, and some type of piety.

But is the journalistic impression accurate? To date there has been precious little sociological research on the topic of "baby-boomer" religiosity (the senior editor of this title has written one of the few extended analyses of the faith of American "boomers"). Moreover, almost none of the literature, to my knowledge, attempts to test generalizations anywhere outside the U.S. It is precisely this need that inspired the work under review.

Although assembled by a trio of well-known American sociologists of religion, this volume is actually the product of a long (and still too rare) transoceanic collaboration among members of the Société Internationale de Sociologie des Religions. In eleven chapters (plus opening and closing statements by the editors), the book brings together the work of fourteen researchers on fourteen countries, which the editors label "the most advanced contemporary nation-states, often described today as in a late- or post-modern phase." (For the record, they are England, Australia, the U.S., Denmark, Iceland, Finland, Norway, Sweden, Germany, the Netherlands, France, Belgium, Italy, and Greece.) What occurs with these cases, the editors maintain, "is suggestive of religion's fate in other modernizing places."

As wide as this selection is, the countries are treated in varying degrees of detail; Belgium, in fact, rates two different entries, while discussion of Iceland, Denmark, Norway, Sweden, and Finland is collapsed into a single chapter on "the Nordic countries." Indeed, many of these contributions, like Eileen Barker's elegant examination of the "gentle, secularizing pluralism" of British religion; or the lucid study of Roman Catholicism in France by Danièle Hervieu-Léger; or Liliane Voyé's too brief venture into the world of small Christian groups in Belgium, would be valuable for their descriptive content alone, even if they were trimmed of all theoretical musings. Nevertheless, some conceptual common ground does emerge.

Most of the authors conclude that, in one way or another, younger believers have rejected the moral absolutes that traditional religious authority imposes, and have reduced accordingly their involvement with the forms of religious activity that "establishments" have to offer. Instead, "boomers" of the modern West have opted in favor of a "new spirituality" that is voluntary, open, emotional, and decidedly subjective — if not wholly solipsistic. Like the collection's Italian contributor, Salvatore Abbruzzese, the majority of the writers find that "the true detachment of this generation is not from . . . religious beliefs but from the institutions representing them; it is not a decline of the sacred as much as a decline of the church's authority as a moral guide."

No matter how exceptional and therefore important, this compilation deserves to be followed by much more research with a rigorously cross-national approach. In the meantime, a moderately priced edition would allow these first fruits of comparative effort to circulate among a broader range of readers everywhere — including, and perhaps especially, students.

Judgment and Sensibility: Religion and Stratification.
By E. Digby Baltzell, Edited and with an introduction by Howard G. Schneiderman. Transaction, 1994. 313 pp. $32.95.

Reviewer: PHILLIP E. HAMMOND, *University of California, Santa Barbara*

No sociologist can claim more of Tocqueville's legacy than E. Digby Baltzell, especially when it comes to Tocqueville's fear of the consequences of democratic equality. Where the French commentator referred to broken links of a chain, each concerned only for itself, Baltzell has his own ways of putting the matter: "John Winthrop would have been horrified with the permissive paradise sought by the Me-people of the 1960s." "John F. Kennedy's entering class at Princeton in 1935 [he later transferred] . . . ranked below the national average in IQ . . . ; the entering class today is at the very top levels of the nation academically. . . . The school for gentlemen has been replaced by a school for preprofessional men and women." "A WASP establishment . . . is no longer possible or desirable today. Some sort of authoritative establishment, however, may be a prerequisite for the ordered freedom which we and the British have enjoyed to a greater extent than any other nation in modern times."

All can agree, then, that "judgment" belongs in this book's title, but many teeth will be set on edge by the "sensibility" part. It is not because Baltzell would restore privilege to a Protestant establishment. He clearly recognizes that, while it may have persisted a few decades longer in culturally pluralistic and tolerant Philadelphia than in Puritan Boston and elsewhere, Protestant monopoly of wealth, club membership, board membership on non-profits, political office, etc., were disappearing phenomena by the 1920s. (Robert Handy, in *A Christian America,* even pegged this period "the second disestablishment.")

Some, however, are bound to be upset by Baltzell's nostalgia — for genealogy, for the gentleman's code in sports, for deference, just to name a few. Those more infected by the counter-culture of the 1960s, or the multiculturalism of today, will be inclined only to rejoice in new-found freedoms; they will be less inclined to remember, for example, the craven responses to Joe McCarthy until the Brahmin Joseph Welch challenged him in Senate hearings with an authority based not just on law but also on "class" — or, if you will, on "fair play," "justice," or "tradition."

All such ideas are easy targets in this so-called postmodern period, perhaps inevitably, given the ease in past decades with which able-bodied, Protestant, heterosexual, wealthy, white males had assumed positions of authority without seriously questioning their right to that authority. But this is not what Baltzell is questioning — what he is questioning is the consequence of overthrowing an American "ancient regime" without attention to its replacement. As he puts it, "Until intellectual leaders . . . come to see . . . that freedom of intellect and liberal democracy are only possible in deferential and hierarchical societies, our forever bickering society will degenerate into the anarchy or tyranny that has marked so much of the third and Marxist worlds."

The various chapters of this collection, dating from 1954 but most more recent, provide ample display of Baltzell's views of this serious issue. The fact that he has no solution — to restoring a "civil religion," for example, or renewing a vibrant and uniform "public opinion," or redefining "self-interest rightly understood" — does not negate his perceptive observation on the collapse of the noblesse oblige ethic that was once part of the American elite class.

Having said all that, however, I must register a negative report as well. The book is 313 pages, of which 6 are an index, 23 are a well-crafted introduction by the editor, Baltzell's admiring former graduate student, 130 are articles or essays of

substance (of which about half are original in this volume), and the remaining 154 are reprinted book reviews, newspaper comments, editorials, etc. In other words, half the book is made up of Baltzell's often trenchant, but brief, observations on the passing scene. There is little, therefore, in the way of sustained argument. Because Baltzell is a graceful and witty writer, this situation is not all bad, of course, but it does make for a book with limited appeal.

Status and Sacredness: A General Theory of Status Relations and an Analysis of Indian Culture.
By Murray Milner Jr. Oxford University Press, 1994. 336 pp. Cloth, $60.00; paper, $24.95

Reviewer: BAM DEV SHARDA, *University of Utah*

The Indian caste system has been the subject of analysis by numerous scholars. The dominant theme in these writings has been to illustrate the unique features of the caste system. Best known among the contemporary western writers on this subject are the French structuralist Louis Dumont, and the American Indologists McKim Marriott and Ronald Inden. They argue that the traditional categories in India are *varnas* or caste clusters ranked by level of purity and should be understood "Indic categories" alone. They, and many other writers (with few exceptions), therefore, argue that the Indian caste system is not directly comparable with other systems of stratification.

In his book, *Status and Sacredness,* Murray Milner, Jr. argues against this and other "orientalist" positions. He has undertaken an extensive analysis of India and argues for a comparative approach. He builds a general theory of status relations from the Indian data, but argues that this theory should be applicable to all societies. For this purpose, he incorporates Weberian theory of status groups with the Durkheimian notions of sacredness. This analysis, he argues, could also offer useful insights into the transition of the modern/developed industrial societies to their postmodern/postindustrial phase.

Many great civilizations (e.g., the Greeks and Romans) made remarkable strides but are extinct now. We only read about them in the history books. Indian civilization, on the other hand, has thrived throughout the ages, despite numerous attacks from outside and within. Why? This resilience of the Indian civilization lies, argues Milner, in the Brahmins' genius for making status the key resource rather than wealth or political power. Convertibility of wealth and power and the problems of legitimacy made the elite in other civilizations the object of discord and hence was a major reason for their downfall. Brahmins in India, on the other hand, used the elements of sacredness to insulate the status resource from the profane world. The profane or material resources were allocated to the other twice-born *varnas*/ caste clusters, whose duty it was to protect and provide for the Brahmin elite. Thus the source of their legitimacy remained in their status.

In describing the social categories in traditional India, Milner argues that *ritual status* was and is the core of Brahmin identity and power. Purity of food and association and sexual asceticism remain their major concerns. However, the economic and political elite (the *Vaishyas* and *Kshatryas)* are more concerned with auspiciousness and the removal of inauspiciousness, which only ritually pure people with special knowledge of astrology can do. Thus, status is the only source of legitimation in India. In an ingenious way, Milner shows how the missing categories of renouncers, untouchables, and bandits become part of the system and how the

status of *Vaisya varnas*/castes, though a part of the traditional social categories, remain ambiguous.

To support his arguments, Milner reviews evidence concerning the continuous cultural prominence of Brahmins, the careful restriction of (status) mobility, the centrality of the concept of *dharma* in daily life, the regulation of eating and marriage arrangements, and the pervasiveness of ritual gift giving. Status (nonmaterial sacred resource), argues Milner, is inalienable, and inexpansible and hence a much more stable resource than wealth or power (the material resources).

Status is a zero-sum resource. If some caste group gains status then the others (higher ones) must lose. In order to limit the transfer of status, ritual norms are prescribed for the higher (privileged) castes, which are protected by elaborately coded "secondary norms." Only members of the privileged strata can learn to "decode" these norms, which include useful rules for enacting associations (e.g., marriages). Thus, the rules of commonsuality and endogamy are devised with care to protect the privileges of higher castes and their identities.

It is understandable that the twice-born castes claim higher status but why should the untouchables, the *shudras* (lower castes), or even the poor Brahmins accept their low nonmaterial or material status? Milner's answer is in his formulations of the Hindu concepts of soteriologies (the means to liberation) and eschatology (about the other worlds) which are conceived as structural reversals of this world. Specifically he argues that: (1) whereas there is inalienability of social status in this world and hence no mobility, *samsara* (literally flow/flux) is a form of social mobility through re-incarnation or transformation; (2) whereas purity and pollution differentiate castes in this world, *moksha* (liberation) becomes the goal of all human beings irrespective of caste; and (3) whereas ascription and inheritance is the key to status in this life, individual *karma* or merit (within the framework of one's *dharma*) is the only way to advance toward liberation. These formulations are the most original and convincing that I have seen in the literature.

I have one small correction and a criticism. On page 16 (last paragraph), the number of *puranas* should be eighteen and not sixteen as reported, perhaps a printing error. My criticism is of Milner's analysis of status. He rightly points out that the most important development in sociology in the last quarter century has been the recognition of non-material resources, first in the notion of human capital (education and skills) and more recently in the notion of symbolic capital (variously referred to as social capital, cultural capital, etc.) in the work of Bourdieu and others. However, both of these notions, Milner believes, ultimately reduce nonmaterial resources to material resources.

Milner's analysis is unique in considering "sacredness [as] a form of status." In the discussion, however, status is reduced to notions of sacredness. This is not at all warranted. In the caste hierarchy Brahmins are universally ranked at the top, but in occupational prestige rankings their traditional occupation (i.e., priest) is not placed that high. This indicates to me that status has two dimensions: ritual status of caste members and secular status provided by their occupational roles. Both can be nonmaterial in nature. The first one is ascribed and inherited whereas the second one may be either achieved or inherited. Thus this one limitation of the book may be noted along with its great importance to the literature.

I believe Milner's book will become very influential in the debate between comparativists and Indologists. However, this time around the Indologists, whether of orientalist or imperialist orientation, will be shooting at a real target rather than at a straw man as they have been doing for so long. In my opinion, Milner has outlined a framework for a genuine synthesis of sociology of religion and stratification that has been neglected in the literature so far. This book is a must for South

Asianists and students of comparative stratification. It will also be useful in graduate theory courses in sociology.

Playing by the Rules.
By John Wilson. Wayne State University Press, 1994. 429 pp. $39.95

Reviewer: ARTHUR T. JOHNSON, *University of Maryland Baltimore County*

Playing by the Rules is a wide-ranging effort to explain the relation between state and society as it affects sport or, more specifically, "the process whereby sport has become a public policy domain." Its guiding thesis is that "policy domains are defined primarily by culture." The recurring theme is to explain sport policy outcomes in terms of "the contradictory forces under which the state must operate in capitalist democracies."

Wilson conceptualizes sport and organizes his chapters in terms of individual, organizational, institutional, and societal levels. He argues that the politics of sport marginalize certain groups and limit the liberty of athletes to exploit their talents. Wilson contends that in the U.S. liberty has been emphasized over equality, resulting in unequal access to leisure opportunity as evidenced by the exclusion of African Americans from professional sports and women from sport in general. At the organizational level, the debate is whether sport organizations should be treated as private or public entities. State action shapes modern sport business and harnesses national values to economic goals in order to ensure viable sport enterprises. At the institutional level, public debate centers on the role of sport in the community and the role government should play in ensuring public access to sport. At the societal level, the state uses sport to promote political assimilation and to symbolize national identity, especially in international competition.

The book's chapters treat barriers to participation, athletic eligibility, anti-trust issues, labor relations, governance of sports leagues and associations, franchise location and movement, mass media, state provision of parks, recreation spaces and athletic facilities, political mobilization and integration, and the role of the U.S. in the international sporting order. Each chapter is primarily a descriptive discussion, based mostly on secondary sources from before 1991 and Congressional committee reports and hearings. Discussion of court cases appears to rely on others' analyses rather than the author's.

As comprehensive as this book attempts to be, it fails to address the one area that would appear to offer the most fruitful examination of Wilson's theoretical claims — that of tax policy toward sports organizations. If, as Wilson claims, sport businessmen have been successful in creating "a state policy toward sport that better protects their investment," tax policy would be a valuable source of evidence to test this thesis. Not one reference is made to taxes, however.

Minor flaws undermine the reader's faith in the author's treatment of his material. The overgeneralizations and trivial errors of fact cause consternation. For example, he says that the modern professional athlete is more likely to be black (in baseball, tennis, golf, hockey, auto and horse racing?); the public has been slow to acknowledge that sports is just a business (perhaps in the 1960s and 1970s, but in the 1980s and 1990s?); and sports and the mass media cannot exist without each other. The claim that NFL owner Robert Irsay accepted the highest bid when he relocated the NFL Colts from Baltimore to Indianapolis is not accurate (Baltimore's bid was actually higher). Ronald Reagan did not portray Knute Rockne in the film, *Knute Rockne — All American* (Pat O'Brien did). Northeastern University is not a public institution.

In sum, this is an overly ambitious text which suffers accordingly. Factual errors, poor copyediting (pages 222 and 223 are misplaced), confused time periods, important omissions, overgeneralization, and at times superficial treatment of the facts detract from the author's analysis. His failure in the concluding chapter to connect the historical description to his early theoretical introduction weakens the book's potential value as a needed source to provide meaning to how sport, the state, and society are related in the U.S.

Engineering Culture: Control and Commitment in a High-Tech Corporation.
By Gideon Kunda. Temple University Press. 1992. 297 pp. Cloth, $34.95.

Reviewer: NANCY DITOMASO, *Rutgers University*

This is a fascinating book for several reasons. It is beautifully written. It is a model of what a book on organizational culture should be (and it recently won the prize for the best book in the Culture Section of the American Sociological Association). It focuses on process and provides the feel of the organization, or in Evelyn Keller's words, the "feeling for the organism." But it is also an anachronism. It is a description of a "super" company that only a few years ago was a model for everyone else, but which has more recently fallen out of favor with almost everyone, for both strategic and technological missteps, and therefore, for also mishandling its own culture. In a recent rating of noteworthy corporations, it was rated last in its industry. Further, this book and this company both appeal to images of corporate culture at a time when such a concept has little currency in the new management fads and fashions. Now the talk is of "jobless" and "boundaryless" firms, and the end of the psychological contract. Loyalty is no longer exchanged for job security. Instead, managers are now encouraging employees to think of themselves as "Me, Inc." (i.e., as temporary employees who should keep their eyes on their own careers), rather than expecting the company to be concerned about such matters. Loyalty and commitment, the essence of the culture metaphor, therefore, are impediments in the paradigm of the "new" organization.

Kunda's purpose is to show through an ethnographic portrayal of everyday experience how culture is "engineered." As he says, he tries to examine the "practical, scientifically informed, self-conscious attempts to design and manage culture," offering an ethnography of "lay ethnographers" in a company whose leaders are committed to the utility of culture as a way to run the business. In this regard, Kunda poses two alternative views of engineered culture, collective "highs" or a new, more personal form of management control. Kunda organizes the description of the company culture around three main ideas: ideology, presentation rituals, and the self and organization. He provides some detail about how the images of the company are produced and made available to employees, the practices by which the culture is made meaningful, and the forms of experience that lead organizational members to define themselves and others in the company. Throughout he contrasts the way the culture enhances self-expression and commitment while also sucking up, burning out, and sometimes keeping out those in the organization's midst.

This book should be read at several levels. It provides a well-crafted ethnographic study of everyday life in a large organization during a time of its strength, and shows how the pressures for solving the conflict between individual and organizational goals affect both the way the organization manages itself and the responses of organizational members. It also, by both its contents and themes, provides a cautionary note for academics who would like to make a difference. The

research for this book was done about a decade ago, and its appearance now is a commentary on how long it takes to do truly good work. But, it appears just as the management winds have shifted, from the frenzy about corporate culture to the equally evangelical fervor about de-bureaucratization, boundaryless and jobless firms, and networks of dense relationships with no center. Perhaps the warnings about the "dark side" of culture have led corporate leaders to abandon it as well, but not because of the "symbolic violence" it engenders. Instead, companies seem to have abandoned the attempt to engineer culture, because of the potential costs of having employees so personally involved and expectant about the company's future. Thus, we can, like Kunda, pose a new dilemma. Is the lack of control and commitment a new form of tyranny or, for some, a new form of freedom and opportunity?

Response to Disaster: Fact Versus Fiction and Its Perpetuation, The Sociology of Disaster.
By Henry W. Fischer, III. University of America, 1994. 160 pp. Cloth, $46.50; paper, $18.50.

Reviewer: RUSSELL R. DYNES, *University of Delaware*

Fischer's short book focuses on a restudy of certain findings from early disaster research, which was unsystematic. Fischer is particularly interested in disaster mythology.

In the early 1950s, the first social science disaster researchers concluded that human beings do not behave in commonly expected ways during disasters: they do not flee in panic, they are not rendered passive, nor do they engage in antisocial behavior. This evidence, however, seems to have had little effect on the beliefs of emergency managers, community officials, and ordinary citizens. These beliefs have come to be called "disaster myths."

Fischer, in a short introductory chapter, provides an initial conceptual discussion as well as a short history of the research field. He then reviews the discrepancy between disaster mythology and research on disaster response, and studies the role of the mass media in perpetuating this mythology, using forty years' worth of disaster stories from a national news magazine. Lastly, he offers an intensive case study of the media coverage of Hurricane Gilbert in 1988 in which he suggests that local TV coverage sought to present information helpful to the community while network TV wanted a "good" story, which meant using more mythology.

Fischer also looks at the issue of developing an effective organizational response to disaster impact, focusing first on the role and training of local emergency managers. He finds that they are more accurately informed about disaster myths than a similar group twenty years before, which underscores the importance of education and training. He then looks at the impact of the media on local emergency operation centers. This case study indicates that a relatively small amount of press criticism evoked significant modification in the behavior of local emergency managers. Finally, Fischer examines the role of experience on the part of the local emergency manager in the creation of mitigation planning.

In his effort to replicate and cumulate such research evidence, Fischer uses surveys, participant observation, field interviews, and content analysis. In addition, he has used aggregate analysis combined with intensive case studies which often illustrate his findings more effectively and dramatically. While the book does not open any new areas of investigation, it strengthens the foundation on which current disaster research rests and that is useful.

Speaking of Life: Horizons of Meaning for Nursing Home Residents.
By Jaber F. Gubrium. Aldine de Gruyter, 1993. 197 pp. Cloth, $43.95; Paper, $22.95.

Reviewer: GEORGE L. MADDOX, *Duke University*

In a notably provocative foreword, David Maines proposes this slim volume as a partial antidote to what he perceives to be a weariness pervading sociology. This weariness is attributed to formula sociology (do a survey, run some statistics, test a model). In this process human agency is neglected and constructed persons replace people. Sociologists study interconnected variables, not persons in contexts. A contemporary student of sociology in an era of secondary analysis of large scale data sets may never meet any subjects face to face or observe personally what people do in various organizational or social contexts. Sociology remains the only academic behavioral and social discipline that has no acknowledged complementary applied component. These are now old complaints among sociology graduate students and some faculty whose hopes that social psychology or a phenomenological perspective would keep people in sociological analysis or bring them back to a more prominent place have not been fulfilled.

Gubrium proposes to do "nursing home ethnography," to give a face to go with a name, to explore the meaning residents perceive in their existence, and to find the possible connections and patterns between the meaning and their earlier lives. In this exploration he studies the narrative accounts of selected residents of nursing homes in response to an invitation to tell their story, to write their autobiographies chapter by chapter. This book, Gubrium writes, takes the "nursing home resident's life narrative as a point of departure for uncovering the subjective meaning of the nursing home experience in relation to the life as a whole." In such accounts one listens for themes and connections of the storyteller's own choosing. Gubrium uses the construct "horizons of meaning" to identify "patterns of narrative linkages each (person) makes with experience in and out of the nursing home."

Fifty-eight residents in several Florida nursing homes were interviewed, 33 of whom had been in residence over a year, a few for more than five years. Twenty-five residents were interviewed multiple times. No detail is given regarding issues of representativeness, reliability or validity of observations, or a theoretical framework for hypothesis or conclusions.

What the reader gets is what Gubrium promised: Very readable edited autobiographical narratives are organized under ten "orientations" such as "worried to death," "making a new home," "spouses" and so on. These observations are not theoretically derived; they are just observed to be there. Hence there is no report of the reliability or validity of the assignment. A relevant contrast in using social biography to predict "successful aging" is Williams and Wirths' *Lives Through the Years* (1965)

In a final chapter, the author does propose some lessons. One is that the lumping together of institutionalized older adults simply as "nursing home residents" is unjustified. Although no developmental theory is apparent in this volume beyond the expectation of lifecourse continuities revealed in the narratives, the storytellers are clearly role makers and role negotiators, not simply role takers. The insights and the sense of self and of efficacy occasionally revealed in some of the stories are notable. In fact, the cognitive intactness of the story tellers is also notable, reminding the reader that the probability of finding such cognitively intact older residents in a nursing home (average age 84) is perhaps in the range of one in two.

The methodological lessons Gubrium proposes involve the importance of giving the cognitively intact respondents free range to tell their stories, to ramble across

their narrative horizons. But the encounters were not freeform. An appendix sketches out an interviewer's guide for the life narrative project. Seeing the residents face to face, up close and personal is perhaps the only way, in the final analysis, to address the currently important issue of assessing the quality of life in long term care settings — which is not, Gubrium argues, equivalent to certifying the technological adequacy of care provided.

The book makes no attempt to estimate at the total quality of life of the residents interviewed and does not systematically characterize the facilities in which these residents live. Gubrium self-consciously avoids formal theory and settles for minimum interpretation and conclusions. The volume is "nursing home ethnography" and fulfills the promise of excellent narrative accounts of the lives of selected residents. Gubrium's next ethnographic account of nursing homes might usefully focus on the life stories of the providers of care, particularly the aides who spend their days interacting with residents.

Painful Inheritance: Health and the New Generation of Fatherless Families.
By Ronald J. Angel and Jacqueline L. Angel. University of Wisconsin Press, 1993. 264 pp. Cloth, $52.00; paper, $19.95.

Reviewer: DULA J. ESPINOSA, *Arizona State University*

Painful Inheritance is groundbreaking because it is one of the first comprehensive examinations of "the impact of minority group status, gender, and poverty on the physical and mental health of children and adult women in fatherless families." The book, which is carefully constructed and thoughtfully presented, reviews and critically analyzes measures of physical, emotional, and social aspects of health like infant mortality and immunization among children and stress and depression among mothers. The book also includes information on Hispanic families (a welcome departure from the more typical black-white only comparisons) and an examination of the effects of childcare on the health of older mostly Hispanic and Black women who are increasingly being called upon to care for their grandchildren.

Chapters 2 and 3 are particularly strong. The magnitude of the problem of "fatherless families" is clearly addressed in chapter 2, which documents the rise of single motherhood among blacks, Hispanics, and whites since 1960. While divorce is a major factor, the number of single teenage mothers is also increasing. The low education and low skills of many of these young mothers inevitably condemn them and their children to a life with little opportunity for social advancement.

The link between national policy and the health of people in single female-headed families is discussed in chapter 3. For example, a table shows how the U.S. differs from other industrialized nations in the types of universal entitlement support it provides to its citizen-families: Great Britain, Norway, France, West Germany, and Sweden, for example, provide family allowance, child allowance, housing allowance, health services, and parental leave. Speculating on possible effects of reductions in social service spending in the U.S., the authors cite the increase in infant mortality that began after cuts in prenatal care in the mid-1980s.

In later chapters, the authors review previous research and present new findings from their own research to show that single mothers generally do "have poorer physical as well as mental health than do married mothers," probably because they are more likely to be poor. This research also shows that black and Hispanic mothers are at an even greater risk for serious health problems than single white mothers.

In sum, *Painful Inheritance* shows that women and children in female-headed families are at a greater health risk than their counterparts in married families. Although the questions raised and answered by this work are important, more research will clarify the links between family structure and health risks. Among the most important of these questions are the health risks of children in "motherless" families, the health risks of children in two-parent families where one parent is abusive, and the potentially offsetting effects of a positive strong male (female) presence in a fatherless (motherless) family. Perhaps the most important question raised by this book is why people in the U.S. are so willing to allow poor women and children to suffer needlessly in order to maintain a relatively low tax base.

Femininities, Masculinities, Sexualities: Freud and Beyond.
By Nancy J. Chodorow. University Press of Kentucky, 1994. 132 pp. Cloth, $20.00; paper, $9.95.

Reviewer: CHRISTINE L. WILLIAMS, *University of Texas at Austin*

Psychoanalytic theory is routinely vilified by sociologists, including many who study gender and sexuality. Uppermost among their reasons for rejecting the theory is the presumed universality of its claims: After all, didn't Freud maintain that every boy experiences castration anxiety and every girl suffers from penis envy? Chodorow offers a compelling response to this criticism in *Femininities, Masculinities, Sexualities.* This book, a collection of three essays originally delivered as lectures at the University of Kentucky, reexamines psychoanalytic theories of gender, sexuality, and love. While Chodorow acknowledges that there are universalizing and normative strands throughout psychoanalytic writings on these topics, she argues that a significant undercurrent of diversity and difference is also present, particularly in the clinical case studies. In fact, her own direct involvement in clinical practice has led Chodorow to abandon most generalizations (psychoanalytic or otherwise) about "men," "women," "heterosexuals," and "homosexuals." These essays convincingly argue that the methods and insights of psychoanalytic theory can remain relevant in our postmodern age of indeterminacy and individual difference.

In the first essay, "Rethinking Freud on Women," Chodorow demonstrates that Freud did not have one all-purpose view of women. Freud's accounts of his women patients and his dealings with women psychoanalysts reveal remarkable tolerance and understanding of a wide range of feminine possibilities. Oddly, the subtlety and variation of these accounts are missing from his grand theory of "normal femininity." Why the difference? Chodorow contends that the femininity described in Freud's theory of female development can best be understood as "what femininity and women mean in the masculine psyche" — in other words, the theory of penis envy is a man's view of what a woman ought to be. As such, it tells us little about women's actual psychical development, but quite a lot about some men's (Freud's?) fantasies and conflicts regarding women.

The second essay, "Heterosexuality as a Compromise Formation," examines the problems with the psychoanalytic account of "object choice" (what most people call sexual orientation). Because Freud considered heterosexuality a natural and inevitable outcome of "normal" gender identity development, he spent little effort explaining how heterosexuality develops from the original bisexuality of infants. In contrast, Freud and his latter-day followers have developed a rich and nuanced account of homosexuality and the so-called perversions. In this essay, Chodorow makes a strong case that there are no clinical or theoretical reasons for treating heterosexuality and homosexuality as dichotomies: she writes that "the sexual

stories, conscious and unconscious fantasies, and transference processes of heterosexuals are as complex and individualized as those of homosexuals."

While Chodorow clearly rejects the privileging of heterosexuality in psychoanalytic theory, she also refuses to consider homosexuality a superior object choice. She suspects that there are probably objective indicators of healthy sexuality, but they have little to do with sexual orientation, *per se*. Thus, instead of trying to figure out how gays and lesbians differ psychologically from straights, Chodorow urges psychoanalytic researchers to develop new ways of distinguishing neurotic symptoms and disorders from passionate and healthy adult sexuality.

The final essay, "Individuality and Difference in How Women and Men Love," argues explicitly against universalistic theories of gender differences. Swimming against the tide of books which maintain some version of the Women-are-from-Venus-and-Men-are-from-Mars theme, Chodorow claims that "men and women love in as many ways as there are men and women." In any individual case, she maintains, gender influences the dynamics of emotional attachment in unique and particular ways.

Throughout these essays, it is apparent that Chodorow's growing familiarity with clinical case studies has shifted her own theoretical focus away from the universal claims about gender and sexuality implicit in Freud (and in some of Chodorow's own earlier work), toward a focus on individuality and diversity. For feminist sociologists, this new synthesis of psychoanalytic theory and postmodernist concerns arrives not a moment too soon.

Take Note

Bailey, Kenneth D. TYPOLOGIES AND TAXONOMIES: AN INTRODUCTION TO CLASSIFICATION TECHNIQUES. *Sage, 1994. 96 pp. Paper, $9.50.* This treatment of classification techniques concentrates on ordering cases in terms of their similarity. Two approaches are discussed — typology and taxonomy. A typology emphasizes the conceptualization of distinct dimensions along which the subject of interest varies. A taxonomy classifies cases empirically using their measured similarity on observed variables. After an overview of each technique the link between the two is explored.

Baxter, Vern K. LABOR AND POLITICS IN THE U.S. POSTAL SERVICE. *Plenum, 1994. 273 pp. Cloth, $39.50.* What were the antecedents and consequences of the reorganization of the U.S. Postal Service. This in-depth study of the country's largest civilian employer addresses the important theoretical questions related to employment, industry restructuring, technology and the political process of change.

Davis, Arthur K. FAREWELL TO EARTH, VOLUME II. *Adamant, Vermont: Adamant Press, 1993. 800 pp. $60.00.* Questioning assumptions about social change and social process, this collection of writings meanders through subjects as diverse as eugenics, industrialization, and mountaineering. A significant portion of the work is dedicated to essays on Thorstein Veblen and a section called "To the Socialist Nation." The latter reviews visits to China, Cuba, and the former USSR.

Jenkins, Philip. USING MURDER: THE SOCIAL CONSTRUCTION OF SERIAL HOMICIDE. *Aldine de Gruyter, 1994, 262 pp. Cloth, $41.95; paper, $20.95.* Jenkins explores the issue of serial murder and its construction into a social problem beginning in the late 1970s. The author examines the social and political climate that has surrounded this topic and how they have shaped the perspectives about serial homicide. He also looks at the role the mass media has played in creating the public sentiment of serial murder as a social problem.

Lee, Raymond M. DOING RESEARCH ON SENSITIVE TOPICS. *Sage, 1993. 248 pp. Paper, $18.95.* In an attempt to guide those engaged in research in sensitive areas, the author explores the potential impact of sensitivity at various stages in the research process — including formulation of the research question, design and implementation of the study, and dissemination of the findings. Additionally, the impact of ethical, political, and legal contexts for research practice is examined.

Prichard, Keith W. and R. McLaren Sawyer (eds.). HANDBOOK OF COLLEGE TEACHING: THEORY AND APPLICATIONS. *Greenwood, 1994. 488 pp. Cloth, $75.00.* An edited collection of works examining the issues confronting university and college professors in the classroom today. Drawing upon the knowledge of teaching faculty, educational specialists and theoreticians, the authors have constructed a handbook that addresses the psychology of learning, different teaching methodologies and theories, and common problems encountered in an undergraduate classroom.

Reed, John Shelton. KICKING BACK: FURTHER DISPATCHES FROM THE SOUTH. *University of Missouri Press, 1995. 180 pp. Cloth, $22.50.* Another humorous and perceptive collection of essays on various aspects of the American South by one of the foremost observers of the Southern scene. Topics that are explored include: state mottoes and mascots; stock car races; a Civil War reenactment; and politicians.

Rubery, Jill and Frank Wilkinson (eds.). EMPLOYER STRATEGY AND THE LABOUR MARKET. *Oxford University Press, 1994. 385 pp. Cloth, $65.00.* The role of employers in determining the quantity and quality of employment has taken on new importance in view of the rapid pace of industrial restructuring and the emergence of new employment policies. Drawing on new data from the Social Change and Economic Life initiative, this book examines how employers' policies impact employment structures and individuals' employment opportunities.

Taylor, Avril. WOMEN DRUG USERS: AN ETHNOGRAPHY OF A FEMALE INJECTING COMMUNITY. *Oxford University Press, 1993. 182 pp. Cloth, $35.00.* Ethnographical account of the lives and experiences of female drug users in Glasgow, Scotland. Taylor explores diverse issues such as how and why these women became drug users, with whom these women socialize, how they support their drug use, and their perceptions of the treatment options available to them.

Zelinsky, Wilbur. EXPLORING THE BELOVED COUNTRY: GEOGRAPHICAL FORAYS INTO AMERICAN SOCIETY AND CULTURE. *University of Iowa Press, 1994. 603 pp. Cloth, $49.95; paper, $22.95.* A marvelous collection of previously published articles by one of our most prolific and inventive cultural geographers treats everything from the spread of classical place names in the 19th century to the geography of Heaven, as revealed in cemetery names. Thought-provoking reading for anyone with the slightest curiosity about American culture and society.

Errata

In Kilbourne, England, and Beron's article "Effects of Individual, Occupational, and Industrial Characteristics on Earnings: Intersections of Race and Gender," published in the June 1994 issue (72:4), Tables 2 and 4 contain some incorrect data. The correct versions of the tables have been reprinted on the following two pages. Readers should note that while substantive conclusions made in the article remain largely unchanged, changes in the tables necessitate some changes in text where specific numbers are mentioned. For a corrected version of the entire article, please contact *Social Forces* at CB #3355, University of North Carolina, Chapel Hill, NC 27599.

TABLE 2: Decomposition of the Gender Gap in Pay, for Whites and Blacks[a]

	Male Mean	Female Mean	Percent Gap Explained Using: Male Slopes	Female Slopes	Weighted Average of Male and Female Slopes
Whites					
Experience (weeks)	342.01	265.07	38.5	29.6	35.1
Education (years)	12.92	12.80	2.6	3.3	2.9
Marital status	.64	.55	2.0	0	1.2
Percent female in occupation	19.52	79.60	3.0	10.0	5.6
Nurturant skill in occupation	-.02	.36	2.4	4.0	3.0
Cognitive skill in occupation	.15	.07	.2	1.1	.4
Industry position	.49	.66	2.6	0	1.6
Blacks					
Experience (weeks)	288.16	255.81	33.5	23.9	29.9
Education (years)	11.14	11.98	-34.6	-43.7	-38.0
Marital status	.47	.42	1.7	.5	1.3
Percent female in occupation	23.70	76.60	5.2	41.7	19.1
Nurturant skill in occupation	-.23	.07	2.5	2.0	2.3
Cognitive skill in occupation	-.70	-.47	-3.3	-1.8	-2.7
Industry position	.50	.63	4.7	.3	3.0

[a] Decomposition assesses percent of the natural log of hourly earnings gap between men and women explained by sex differences in means on each variable. Slopes used are those in Table 1.

TABLE 4: Decomposition of Race Gap in Pay, for Men and Women[a]

	White Mean	Black Mean	Percent Gap Explained Using: White Slopes	Black Slopes	Weighted Average of White and Black Slopes
Women					
Experience (weeks)	265.07	255.81	16.8	16.8	16.8
Education (years)	12.80	11.98	107.0	104.7	106.5
Marital status (married = 1)	.55	.42	.1	3.0	.7
Percent female in occupation	70.96	70.66	-0.2	-0.7	-0.3
Nurturant skill in occupation	0.36	0.07	-14.5	-4.7	-12.3
Cognitive skill in occupation	0.07	-0.47	33.7	10.1	28.3
Industry position (periphery = 1)	0.66	0.63	.1	.1	0
Men					
Experience (weeks)	342.01	288.13	38.9	41.9	39.6
Education (years)	12.92	11.14	56.3	54.9	56.0
Marital status (married = 1)	0.64	0.47	5.4	4.4	5.2
Percent female in occupation	19.52	23.70	0.3	0.3	0.3
Nurturant skill in occupation	-0.02	-0.23	-1.9	-1.3	-1.7
Cognitive skill in occupation	0.15	-0.70	4.4	9.2	5.5
Industry position (periphery = 1)	0.49	0.50	0.2	0.3	0.2

[a] Decomposition assesses percent of the natural log of hourly earnings gap between blacks and whites explained by race differences in means on each variable. Slopes used are those in Table 1.

SOCIAL FORCES

University of North Carolina Press
P.O. Box 2288
Chapel Hill, NC 27515-2288

Second class postage paid at Chapel Hill, N.C. and additional post offices

RETURN POSTAGE GUARANTEED

SOCIAL FORCES

AN INTERNATIONAL JOURNAL OF SOCIAL RESEARCH
ASSOCIATED WITH THE SOUTHERN SOCIOLOGICAL SOCIETY

VOLUME 74: NUMBER 2 DECEMBER 1995

SF AUTHORS' GUIDE

Revised February 1994

IN GENERAL

1. Mail manuscripts (5 copies) to: **Social Forces**, Department of Sociology, Hamilton Hall, University of North Carolina, Chapel Hill, NC 27599-3210.
2. The paper must not be submitted to another journal while under review by SF.
3. Enclose a $15 processing fee. Papers are not reviewed until this fee is paid. The fee is waived for papers solely by currently enrolled students.
4. Manuscripts are not returned to authors.
5. Authors transfer copyright to SF. Permission fees for reprint of articles are split between authors and journal.

PREPARATION OF MANUSCRIPTS

1. For review, papers may be submitted in the form used by SF, American Sociological Association or other major sociological journals. Authors of accepted manuscripts are responsible for putting them into SF form.
2. **Format, Organization, and Style** Submit 5 clearly legible copies on 8 ½ by 11 good quality white paper. Double space typing throughout. For an accepted paper being revised to fit SF style, leave at least 1" margins; do not justify margins if using a word processor.

i. **Order** Organize the manuscript in this order: cover page; abstract; text; endnotes; references; tables; figures.

ii. **Cover Page** Give title; author(s); affiliation(s); and a footnote (*) indicating name and address of the author to whom requests for offprints or other correspondence should be sent ("Direct correspondence to ___") and acknowledgment (if any) of financial or other assistance.

iii. **Abstract** On a separate page, preceding the text, write a summary, 125 or fewer words (70 or fewer for a Research Note).

iv. **Endnotes** Use only for substantive comments, bearing on content. Number consecutively from 1, double space, and append on a separate page.

v. **Tables** Type each table on a separate page. Insert a location note — "Table 1 about here" — in the text. For substantive footnotes relating to a table or figure use lowercase letters. Use tabs to separate columns.

vi. **Figures** Must be artist-drawn, camera ready. Do not send original copy with a manuscript submitted for review. For review, figures may be roughly drawn but they must be neat and legible. For footnotes to figures, use same symbols as for footnotes to tables.

vii. **Mathematical Symbols** (and others likely to be obscure to compositor) Clarify with encircled words written in the margin.

viii. **References in Text** Indicate sources as illustrated below:

1. when author's name is in text — Lipset (1960); when author's name is not in text (Lipset 1960)
2. use page numbers only for direct quotations or specific notes or table — (Braudel 1969:213-14)
3. for more than 3 authors use "et al."
4. with more than 1 reference to an author in the same year, distinguish them by the use of letters (a,b,c) with year of publication (1975a)
5. earlier publication should precede later publication in brackets with parentheses (Tocqueville [1835] 1956)
6. enclose a series of references — in alphabetical order — in parentheses, separated by semicolons (e.g., Adler 1975; Adler & Simon 1979; Anderson, Chiricos & Waldo 1977; Bernstein et al. 1977, Chesney-Ling 1973a, 1973b).

ix. **References Following Endnotes** List authors alphabetically, by surname. For authors with more than one work cited, list works earliest to latest. For articles, next give title of article (caps and lower case), name of journal, volume number, and pagination. For books and monographs, give title, followed by publisher.

x. **Format of References**

Please spell out the first names of all authors and editors, unless they use only their initials or a first initial and a middle name in the source cited (e.g., Paul Radin, T.S. Eliot, and J. Owen Dorsey).

Elder, Glen H. 1975. "Age Differentiation and the Life Course." Pp. 165-90 in *Annual Review of Sociology*. Vol. 1, edited by Alex Inkeles, James Coleman, and Neil Smelser. Annual Reviews.

Myrdal, Gunnar. [1944] 1962. *An American Dilemma*. Harper & Row.

Ritzer, George. 1975a. *Sociology: A Multiple Paradigm Science*. Allyn & Bacon.

____. 1975b. "Sociology: A Multiple Paradigm Science." *American Sociologist* 10:156-67.

Computer Files Authors analyzing data available in this format should cite their source, indicating producers and distributors as in these examples:

Louis Harris and Associates. 1975. Harris 1975 Nuclear Power Survey #2515 [computer file]. New York: Louis Harris and Associates [producer]. Chapel Hill: Louis Harris Data Center, University of North Carolina [distributor].

U.S. Bureau of the Census. 1979. 1970 Census of Population and Housing, Fourth Count Population Summary Tape [computer file]. Washington: U.S. Bureau of the Census [producer]. Rosslyn, Va.: DUALabs [distributor].

SOCIAL FORCES

SEPTEMBER 1995, VOLUME 74, NUMBER 2

ISSN 0037-7732 Library of Congress Catalog Number 24-31023

Second class postage paid at Chapel Hill, N.C. 27514 and additional mailing offices. *Postmaster: Send 3579 to Social Forces, UNC Press, P.O. Box 2288, Chapel Hill, NC 27515*

Claims for undelivered copies must be made within the month following the regular month of publication. The publisher will supply missing copies when losses have been sustained in transit and when the reserve stock will permit.

Social Forces is issued quarterly: September, December, March, June.

Social *Forces* is published by the University of North Carolina Press for the Department of Sociology, with the assistance of the Institute for Research in Social Science, University of North Carolina at Chapel Hill.

Rev. 7/3/95

INFORMATION FOR *SF* SUBSCRIBERS

For Volume 74

Social Forces is issued quarterly: September, December, March, June.

Subscriptions are accepted by the volume only, i.e., beginning in September and ending with the June number. Subscriptions entered after the appearance of the first issue in a volume should specify whether the subscriber wants the current volume (including back numbers) or wants to begin the subscription with the following volume.

Subscription Rates

INDIVIDUALS - U.S.		INDIVIDUALS - OUTSIDE U.S.	
Regular:		Regular:	
1 year:	$30	1 year:	US $38
3 Years:	$85	3 Years:	US $110
ASA:		BSA, CSAA, ASA (non-U.S.):	
1 year:	$25	1 year:	US $33
3 years:	$70	3 years:	US $95
Students:		Students:	
1 year:	$12	1 year:	US $20

INSTITUTIONS - U.S.		INSTITUTIONS - OUTSIDE U.S.	
1 year:	$52	1 year:	US $60
3 years:	$145	3 years:	US $170

Single Copies - U.S.
US $13

***Social Forces* Cumulative Indexes**
Volumes 1-50 (1922-1972): US $12
Volumes 51-58 (1972-1980): US $12

(For single copies and the Index add US $1.50 for mailing outside the U.S.)

Back Numbers

Reprinted Volumes 1 through 40 are available from the Johnson Reprint Corporation, 111 Fifth Avenue, New York, New York 10003. Back numbers of Volume 41 and subsequent issues are available from the managing editor of *Social Forces*, Institute for Research in Social Science, Manning Hall CB #3355, University of North Carolina, Chapel Hill, North Carolina 27599. The four most recent back volumes are available from The University of North Carolina Press. Volumes are available in microfilm from University Microfilms, 300 North Zeeb Road, Ann Arbor, Michigan 48106.

Abstracts and Indexes

The articles in *Social Forces* are indexed or abstracted in *A.B.C. Political Science & Government, Abstracts in Anthropology, Academic Index, Adolescent Mental Health Abstracts, American Bibliography of Slavic & East European Studies, Applied Social Sciences Index & Abstracts, Automatic Subject Citation Alert, Chicano Periodical Index, Communication Abstracts, Criminal Justice Abstracts, Current Contents, Current Index to Journals in Education, Current Literature in Family Planning, Education Administration Abstracts, Excerpta Indonesica, Family Studies Abstracts, Geographical Abstracts, Historical Abstracts, Leisure, Recreation & Tourism Abstracts, Linguistics & Language Behaviour Abstracts, Middle East: Abstracts & Index, PAIS: Public Affairs Information Service, Population Index, Psychological Abstracts, Social Work Research & Abstracts, Social Sciences Index, Social Sciences Abstracts, Social Science Citation Index, Sociological Abstracts, Studies on Women Abstracts, Urban Affairs Abstracts, Urban Studies Abstracts, and World Agricultural Economics & Rural Sociology Abstracts, International Bibliography of Periodical Literature*, and the book reviews in *Book Review Index.*

Change of Address

Notice of change of address should reach the publisher at least 6 weeks before publication of the next number. Please include both old and new addresses and zip codes.

Communications regarding subscriptions should be addressed to: Social Forces Subscriptions, The University of North Carolina Press, Box 2288, Chapel Hill, North Carolina 27515

Editorial and business communications should be addressed to: Social Forces, 168 Hamilton Hall, University of North Carolina, Chapel Hill, North Carolina 27599-3210.

SOCIAL FORCES

DECEMBER 1995, VOLUME 74, NUMBER 2

Contents

BOOK REVIEWS

The Thomas Theorem and The Matthew Effect*

ROBERT K. MERTON, *Columbia University and Russell Sage Foundation*

Eponymy in science is the practice of affixing the names of scientists to what they have discovered or are believed to have discovered,[1] as with Boyle's Law, Halley's comet, Fourier's transform, Planck's constant, the Rorschach test, the Gini coefficient, and the Thomas theorem.

This article can be read from various sociological perspectives.[2] Most specifically, it records an epistolary episode in the sociointellectual history of what has

[1] The definition of *eponymy* includes the cautionary phrase, "or are *believed* to have discovered," in order to take due note of "Stigler's Law of Eponymy" which in its strongest and "simplest form is this: 'No scientific discovery is named after its original discoverer'" (Stigler 1980). Stigler's study of what is generally known as "the normal distribution" or "the Gaussian distribution" as a case in point of his ironically self-exemplifying eponymous law is based in part on its eponymous appearance in 80 textbooks of statistics, from 1816 to 1976.

[2] As will become evident, this discursive composite of archival documents, biography of a sociological idea, and analysis of social mechanisms involved in the diffusion of that idea departs from the tidy format that has come to be prescribed for the scientific paper. This is by design and with the indulgent consent of the editor of *Social Forces*. But then, that only speaks for a continuing largeness of spirit of its editorial policy which, back in 1934, allowed the ironic phrase "enlightened Boojum of Positivism" (with its allusion to Lewis Carroll's immortal *The Hunting of the Snark*) to appear in my very first article, published in this journal better than 60 years ago.

** I am indebted, once again, to Harriet Zuckerman, Robert C. Merton, Cynthia Fuchs Epstein, David L. Sills, and Stephen M. Stigler for vetting a manuscript, to Jennifer Lee and Maritsa Poros for research assistance, and to Eugene Garfield for aid of other kinds.*
Direct correspondence to Robert K. Merton, East Gallery, Low Memorial Library, Columbia University, New York, NY 10027.

 Social Forces, December 1995, 74(2):379-424

come to be known as "the Thomas theorem":[3] "if men define situations as real, they are real in their consequences" (Thomas & Thomas 1928:572). More generally, that episode provides a strategic research site for examining certain substantive and methodological problems in the sociology of science. From the methodological perspective, it provides a prime example of the basic if presumably obvious precept that it is one thing to establish a phenomenon (i.e., show that something is empirically the case) and quite another to explain it. Obvious this may be but, as we shall see, the two are nevertheless easily (and, I believe, often) conflated. The episode also exhibits the risks involved in reductionist, single-factor explanations of a concrete social phenomenon (which becomes even more marked in *ex post* than in *ex ante* explanations). Finally, the episode provides an apparent instance of how sociocultural contexts of science and scholarship — in this case, the belated thrust of the civil rights movement toward equity for women — can make for an exclusive and premature interpretation that a particular cognitive phenomenon is sexist.

The cognitive phenomenon in this case consists of sociologists' frequent ascription of the Thomas theorem solely to W.I. Thomas rather than to both W.I. and Dorothy Swaine Thomas. I should emphasize from the start, however, that there is only incidental interest here in trying to adjudicate proprietary claims to this basic sociological idea, although the introduction of private archival materials may contribute to that result. Rather, the widespread accreditation of the Thomas theorem to W.I. Thomas alone holds interest for us here principally as a specific instance of a generic phenomenon in the reward-system of science and scholarship — what can be conveniently described as "the partial citation phenomenon," thus substituting four words for the approximately 20 words of its definition, i.e., the widespread accrediting by scientific and scholarly peers of an (actually or apparently) joint contribution to only a subset of the collaborators.[4]

Some forms of the partial citation phenomenon almost provide their own, intuitively evident, explanations. The partial citation of coauthors is hardly problematical, of course, for the rapidly increasing number of scientific articles,

---

[3] As I have noted elsewhere (Merton 1984:282), the designation Thomas *theorem* "does not, of course, adopt the term *theorem* in the strict mathematical sense (as, say, with the binomial theorem). It refers, rather, to an idea that is being proposed or accepted as sound, consequential, and empirically relevant." In proposing the Thomas eponym for both mnemonic and commemorative purposes (Merton [1942] 1973:273), I had fastened on the term *theorem* rather than such less formidable terms as *dictum, maxim, proposition*, or *aphorism* in order to convey my sense that this was "probably the single most consequential sentence ever put in print by an American sociologist" (Merton 1976:174). In any case, the word *theorem* was rhetorically employed in the same broad sense that had the mathematically minded Hobbes referring in *The Leviathan* to "general rules, called theorems or aphorisms." (Reading this note, the mathematical statistician, Stephen M. Stigler, reminds me that the redoubtable seventeenth-century mathematician John Wallis — he of the Wallis theorem — had ample cause to destroy the pretensions of the mathematically minded Hobbes to being an actual mathematician.)

[4] The word *citation* in the term *partial citation phenomenon* is to be construed broadly; not only as a formal reference in a note or bibliography but as any mode, including eponyms, of referring to previous scientific or scholarly work.

chiefly in the physical and biological sciences, with large numbers — at the extremes, literally hundreds — of listed coauthors (Science Watch 1995:1-2; Zuckerman 1968). Nor is it greatly problematical in cases of works by, say, four or five authors where the citation pattern of "et al." has long since evolved in the world of scholarship to serve the shared convenience of publishers, editors, senior authors, and readers (if not the neglected collaborators). Nor is it problematical that citations of large-scale research reports often have the institutionalized senior investigator alone eponymized, as with The Kinsey Report or The Coleman Report. But the partial citation phenomenon is surely problematical in the limiting case, as with the book in which the Thomas theorem first appeared, when only one of two authors is regularly cited.

The equitable peer ascription of contributions is no minor matter in the social institution of science which has evolved a reward system that consists basically in rewarding scientists by having knowledgeable peers grant them public recognition for their distinctive contributions. All other rewards flow from it.[5] And so it is that peers will experience and sometimes publicly express strong moral sentiments when they have reason to believe — as in the case of the Thomas theorem some have felt there is reason to believe — that the norm of rightful accreditation has been violated by systematically biassed ascriptions in the pertinent community of scientists. For like all other social institutions, the institution of science has its (partly manifest, partly latent) normative framework, one that includes the norm of *equity* in peer recognition of contributions.

Along with letters exchanged some time ago between Dorothy Swaine Thomas and myself, the core archival materials in this study of the allocation of credit for the Thomas theorem consist principally of a more recent exchange of letters between another pair of collegial sociologists which has one of them interpreting this instance of the partial citation phenomenon as a "piece of (dare I call it?) institutionalized sexism . . . in the new era." These letters — one of them, crisp and pointed by a social scientist of amply merited international fame; the other by myself, and replete with documentary exhibits running to a good-sized article in its own right — will be quoted verbatim.[6] As the more voluble member of that epistolary pair, I shall subject that exchange to analysis in terms of patterns in the growth and transmission of knowledge that I have been investigating over the years: such patterns as "establishing the phenomenon" and the use of "strategic research sites" (Merton 1987); "the retroactive effect" in perceiving or imagining adumbrations and anticipations of ideas in

[5]For the paradigm which maintains that the institutional dynamics of science derives from the interaction of its "reward system" and its "normative structure," see Part IV of Merton 1973.

[6]An apt procedure, one would think, since the sociological analysis of verbatim letters was introduced by W.I. Thomas and Florian Znaniecki; though, to be sure, on a rather larger scale running from page 217 to page 1114 of their classic work in five volumes, *The Polish Peasant in Europe and America* ([1918-20] 1927). For a critical examination of methodological problems involved in the use of letters in *The Polish Peasant*, see Blumer 1939:29-39.

the history of thought;[7] patterns of primary and derivative or serial diffusion of knowledge (Merton 1995); "socially organized skepticism" in the domain of science and scholarship (Merton [1942] 1973:267-78,339-40); the place of eponyms (such as the Thomas theorem) and of other forms of peer citation in the reward system of science (Merton [1942] 1973:273-74; [1957] 1973:297-302; 1988:619-23); "oral publication" as distinct from publication in print (Merton 1980); "multiples" or independent multiple discoveries and inventions (Merton 1973; Ogburn & Thomas 1922) and the emergence of citation analysis (Garfield 1955; Garfield, Sher & Torpie 1964). Above all else, however, the partial-citation phenomenon of the Thomas theorem will be analyzed in terms of the Matthew effect (Merton & Zuckerman [1968] 1973;1988).

This epistolary exchange which involves contesting interpretations of this one case of the partial citation phenomenon also raises normative questions about the peer recognition of scientific and scholarly contributions that provides the ultimate social and moral, if not the legal, basis of intellectual property in science (Zuckerman 1988a:526-27).

(As we have just seen in preview, such analysis in terms of one's enduring thematic interests in the sociology of science is bound, alas, to entail an intemperate abundance of self-citation. It is some consolation to note, however, that disciplined citation analyses do not mistake self-citation as evidence of the peer recognition that is the ultimate coin of the domain of science and scholarship.)

Diffusion of the Thomas Theorem

Like other important sociological ideas, the Thomas theorem has had its adumbrations and partial anticipations. Recognition of the subjective component in human action has had a long history in sociological thought and a far longer history before we sociologists arrived on the historical scene. Among the Ancients, we need only recall Epictetus (*The Encheiridion* [c. 110 A.D.] 1926-28: II, 487, §5), stating that "What disturbs and alarms man are not actions, but opinions and fancies about actions." And if one has not had occasion to read Epictetus recently, his aphorism (theorem?) may be recalled from the still enduring eighteenth-century masterwork, *Tristram Shandy*, where Laurence Sterne quotes it — in Greek of course — on the title page. And among the Moderns, there is Schopenhauer ([1851] 1974:I, 326) also echoing Epictetus[8] as he observes that "it is not what things objectively and actually are, but what they are for us and in our way of looking at them that makes us happy or unhappy." Be it said, however, that, as is generally the case with the *ex post*

[7] "Anticipations" refer to earlier ideas, formulations or findings that overlap later ones but do not focus upon and draw the same implications from them; "adumbrations," to earlier formulations or findings that, quite literally, foreshadowed later ones but only dimly and vaguely. On these patterns in the history of thought, see Merton [1949, 1957] 1968:13-25.

[8] Typically still for his time and place, Schopenhauer of course also echoed Epictetus in the original Greek and the derivative Latin rather than in his own native German.

spotting of adumbrations or anticipations, my having noticed those anticipatory passages is probably just another instance of the "retroactive effect" (in which undeveloped ideas that have remained in oblivion are later brought into the limelight only because new and clearly formulated ideas sensitize us to earlier, typically less developed, and previously ignored, versions).[9] The chances are that I would not have taken note of those aphorisms (theorems?) by Epictetus and Schopenhauer — as quoted even in the various editions of *Tristram Shandy* on my bookshelves — had it not been for a half-century of close familiarity with the Thomas theorem.

Along with these venerable anticipations, the theorem also had a striking contemporary and neighboring version. In his lecture course at the University of Chicago, W.I. Thomas's colleague George H. Mead had observed in distinctly sociological terms that "If a thing is not recognized as true, then it does not function as true in the community." But the Thomas theorem and the Mead theorem experienced notably different cognitive fates. In virtually self-exemplifying style, the Mead theorem dropped into permanent oblivion even after its posthumous transition from "oral publication" in lectures to publication in print (Mead 1936:29). Not so with the Thomas theorem.

True, no notice of the theorem was taken in any of the reviews of *The Child in America* that appeared in the three principal American sociological journals of the time — the *American Journal of Sociology* (which was then not only produced and edited at the University of Chicago but was also the official journal of the American Sociological Society), *Social Forces* (at the University of North Carolina), and *Sociology and Social Research* (at the University of Southern California). But soon afterward, as we shall have occasion to see in detail, Kimball Young, the prolific author of textbooks in social psychology and sociology then at the University of Wisconsin, gave the theorem special notice by selecting it as an epigraph for chapters in two successive textbooks; one, in his widely adopted *Social Psychology* (Young 1930:397) and the other, in his edited volume, *Social Attitudes* (Young 1931:100). The Young epigraphs evidently become an early conduit for diffusion of the quoted sentence. At any rate, the very next year, the omnivorous sociologist Howard [P.] Becker, was interpolating the sentence twice in his amplified and Americanized edition of Leopold von Wiese's *Allgemeine Soziologie* ([1924] 1932:34,79) and faithfully reporting that it was being quoted from Young's *Social Psychology*.

As is often the way with the genealogy of ideas, various types of errors began to intrude with enlarged diffusion. Omnivorous reader though he was, Becker managed to commit a triple error when a dozen or so years later he ostensibly located the Thomas sentence on page 79 of the condensed one-volume edition of the Thomas and Znaniecki masterwork, *The Polish Peasant in Europe and America* ([1918] 1927), where it is not to be found. Not content with this mishap, Becker advanced the extraordinary conclusion that this nonexistent ghost appearance was "in content at least . . . probably Znaniecki's," this even

---

[9]"Undeveloped ideas" inasmuch as these earlier adumbrations or anticipations were not singled out, elucidated, or followed up by further theoretical or empirical inquiry, either by their original authors or by others.

though the concept of "definition of the situation" had appeared before *The Polish Peasant* and then endured continuously as basic in Thomas's conceptual framework but did not turn up at all, Polish colleagues assure me, in Znaniecki's early work and surely not in his later work in English.[10]

The Barnard sociologist Willard Waller (1938:20) introduced an error of another kind in the course of diffusion. He misquoted the Thomas dictum, understandably without citing its specific source, thus: "As Thomas has put it, 'If *people* define *things* as real, they are real in their consequences.'" In our own time of acute sensitivity to deliberately or unwittingly gendered terms, one is at first tempted to interpret Waller's substitution of *people* for *men* as a quiet but deliberate effort to de-gender the theorem. However, as one notices the further substitution of *things* for *situations*, it seems more likely that the de-gendering was the inadvertent result of a faulty memory.[11]

In the same year as Waller's would-be quotation, an effort was made to pinpoint the lucid and elegant character of the formulation by describing it as "W.I. Thomas's sociological theorem" (Merton 1938:333). Like Young's quotations of the theorem, this almost casual allusion made no precise reference to the original source of what soon became abbreviated as "the Thomas theorem." The long-standing and largely continuing absence of such specific citations and texts persuade me that the widely neglected paragraphs which culminated in the Thomas theorem both require and merit repetition here. We note that the theorem caps the methodological case being made for use of

> The behavior document (case study, life-record, psychoanalytic confession) [which] represents a continuity of experience in life situations . . .
>
> [E]ven the highly subjective record has a value for behavior study. A document prepared by one compensating for a feeling of inferiority or elaborating a delusion of persecution is as far as possible from objective reality, but the subject's view of the situation, how he regards it, may be the most important element for interpretation. For his immediate behavior is closely related to his *definition of the situation*, which may be in terms of objective reality or in terms of a subjective appreciation — 'as if' it were so. Very often it is the wide discrepancy between the situation as it seems to others and the situation as it seems to the individual that brings about the overt behavior difficulty. To take an extreme example, the warden of Dannemora prison recently refused to honor the order of the court to send an inmate outside the prison walls for some specific purpose. He excused himself on the ground that the man was too dangerous. He had killed several persons who had the unfortunate habit of talking to themselves on the street. From the movement of their lips he imagined that they were calling him vile names, and he behaved as if this were true. *If men define situations as real, they are real in their consequences* (Thomas & Thomas 1928:572; italics added).

As we see, this essentially methodological observation draws upon the basic substantive concept of "defining the situation" which Thomas (1923:42-43;

---

[10]This statement will surely not be taken to detract from Znaniecki's major contributions to sociology which, however, were of quite other kinds than that represented by the theorem. My own appreciation of those contributions is summarized in Merton 1983.

[11]On the phenomenon of unintended gendering and de-gendering of language, see Merton, "De-Gendering 'Man of Science,'" 1996.

1929:1-13) plainly regarded as his most significant contribution to the socio-psychological understanding of the formation of social personality and character. Seen in its immediate context, the memorable theorem turns out to be a generalization of a specimen of paranoid behavior. (As we shall also see in due course, Thomas largely confined himself to reiterating this example the only two other times he himself quoted the theorem, a singular circumstance that raises an obvious problem for future exploration).

The longtime absence of a correct reference to the source of the theorem in sociological writings quoting it also led me to conclude some time ago that it had become known to American sociologists and their students largely if not entirely through secondary discussions in print rather than through their having read the original text. This assumption was reinforced by inquiries over the years from colleagues near and far which, we shall see, asked for the exact source of what had been described and analyzed as "the Thomas theorem" in "The Self-Fulfilling Prophecy" (Merton 1948). How such secondary diffusion came to be and, in particular, how social mechanisms of *initial* diffusion operated through such pathways as sociocognitive networks will be examined in the last part of this article as we explain sociologically the early absence of precise citations to the theorem and how that in turn helped produce the partial citation phenomenon.

That once conjectural assumption that the theorem had largely become known through secondary sources has now been empirically confirmed in a study by R.S. Smith (1993) of "well over 100 introductory textbooks" of sociology which found only one of the 40 texts that quoted or paraphrased the theorem citing its source, replete with page number, while an unspecified number of authors actually cited "The Self-Fulfilling Prophecy" as a mediating source. Smith also found that most of the textbooks attributed the theorem to W.I. Thomas alone and the few that referred to the book in which the theorem appeared generally failed to cite Dorothy Swaine Thomas as its second author. These empirical findings led Smith to a type of judgment which we have noted is deeply imbedded in the normative framework of science about equity in scholarly attributions:

> [S]ince it [the theorem] appears in a co-authored work, and no particular author is singled out as having written Chapter XIII [in which the theorem appears], it seems reasonable to suppose either author could have written this phrase. Consequently, *unless there is compelling evidence to the contrary,* it would seem proper scholarly practice to attribute these words to both W.I. Thomas and Dorothy Swaine Thomas (Smith 1993; I have inserted italics to underscore the scholarly care with which Smith allows for the possibility of countervailing evidence).

Having reported his empirical findings and having arrived at his contingent normative judgment, Smith goes on to propose possible explanations of this

> widespread practice of excluding Dorothy Swaine Thomas in referencing the theorem. This might be explained by poor scholarship on an individual level, although several of those involved are nationally and even internationally known authors. It can also be explained in terms of a structural issue — the genderization of the discipline as part of the process of professionalization. By not citing Dorothy Swaine Thomas these authors help sustain a view of sociology as historically a male domain (Smith 1993).

However, in mounting this crisp study of "Dorothy Swaine Thomas and the 'Thomas Theorem'" Professor Smith could not possibly have known that both his normative observation about the proper citation of coauthored work and his structural hypothesis of what might have led to the almost exclusive ascription of the eponymous theorem to the male Thomas at the expense of the female Thomas had been independently and concisely stated, without benefit of systematic empirical study, in a letter written some five years before by a social scientist of amply merited international fame.

The Imputation of "Institutionalized Sexism"

That letter, addressed not to me but to David Sills, with whom I was then collaborating on the volume *Social Science Quotations* (Sills & Merton 1991), holds varied historical, methodological, and sociological interest as it briefly describes a search for "the original source of the Thomas theorem." It bears not only upon the early history of this important sociological idea but also upon the methodological requirement of putting to empirical test *ex post* hypotheses evoked by data that are, of course, initially congruent with them. Moreover, as I have hinted, the plainly informal letter and my lengthy thematic responses to it exemplify various patterns in the transmission of knowledge that I had been exploring for some time. Not least in point, the letter culminates in the composite normative and explanatory conclusion that ascribing the theorem to W.I. Thomas alone amounts to sexist eponymy.

Here, then, is the short evocative letter in its verbatim entirety.

August 16, 1988

David Sills
Social Science Quotations
111 Eighth Ave., Suite 1503
New York, NY 10011

Dear David,

I should have the X . . . and Y . . . citations to you shortly. In the meantime, I bring to your attention an interesting footnote in the sociology of sociology. About eight months ago, I was desperately looking for the original source of the Thomas theorem. I could find it nowhere, except in Merton who used it without a footnote. I finally was about to resign myself to using as my citation: Cited in Merton, etc. Then your little brochure, "Social Science Quotations: Guidelines" arrived, and lo and behold, on p. 15, as one of your examples, you use the Thomas theorem, with a proper citation.

Well trained as I am in scholarly skepticism of sources, I took out the book from the library. And lo and behold, you are in error, an error which not only you but Merton and indeed the entire U.S. (world?) scholarly community have made consistently. For the book was <u>not</u> written by W.I. Thomas, but by William I. and Dorothy Swaine Thomas. Nothing in the book indicates that he wrote some chapters and she others. They are joint authors. And, at least in Social Science Quotations, this piece of (dare I call it?) "institutionalized sexism" should not be perpetuated.

Yours in the new era,

[I anonymize the author of this informal note as "Skeptical Social Scientist" (hereafter SSS)]

Patterns in the Transmission of Scientific Knowledge

Although this letter was addressed to David Sills, I undertook to respond to it since I was the one who had long ago elected to attribute the theorem exclusively to W.I. Thomas and since, as coeditor of *Social Science Quotations*, I had prepared the entry on the theorem as a specimen quotation in the "Guidelines" for contributors to that volume. I soon found myself resonating to the considerable array of cognitive and normative ideas and problems packed into the brief but complex letter that exemplified diverse patterns in the normative framework of science and the transmission of scientific knowledge. Herewith, then, a few of those patterns.

PATTERNS OF KNOWLEDGE DIFFUSION

To begin with, it will be noticed that SSS matter-of-factly adopts the eponym "Thomas theorem" which had been introduced a half-century before to signal the assessed importance of the reverberating Thomas sentence. It is unlikely, however, that SSS had come upon that eponym where it first appeared since that was in an article published long ago (Merton 1938) in the *Philosophy of Science*, a journal presumably not often read by social scientists of a much later time. Nor is it likely that this same social scientist writing in the late 1980s would have come upon the eponym in 1948 when the theorem was being analyzed and distinguished from the related concept of "The Self-Fulfilling Prophecy" in *The Antioch Review*, another journal not notable for its social science readership. I am inclined to believe, therefore, that like those others who had let me know that this was the case for them, SSS had come upon the eponym in one or another of the three editions of *Social Theory and Social Structure* (Merton 1949, 1957, 1968) which had included both those articles. In accord with the R.S. Smith (1993, 1995) finding that this book had become something of a conduit for the Thomas sentence, here is SSS "desperately looking for the original source of the Thomas theorem" and about to adopt what some might describe as the honorable but, for many, the also unappetizing expedient of quoting from a mediating source; to wit: "Cited in Merton, etc." (just as we have seen Howard P. Becker do some 60 years before in quoting the theorem via Kimball Young). As indicated by SSS and others, this citational expedient had resulted from the continued absence of any precise reference to the original source in all three editions of *Social Theory and Social Structure*.[12] Fortunately, SSS informs us, *Guidelines for Contributors to Social Science Quotations* came along in the nick of time to provide "a proper citation."[13]

---

[12]As we shall see from archival evidence yet to be examined, both adventitious and theoretical reasons made for this studied failure to supply the originally well-known reference to the exact source of the Thomas theorem in *The Child in America*.

[13]This, then, may have been the first anticipatory case, even before publication, in which *Social Science Quotations* fulfilled a manifest function that was described this way: "Of obvious use to readers coming upon quotations new to them, exact references may also prove useful for swiftly locating the more familiar quotations. By leading readers back to the sources, such detailed references can help them place even extended quotations in their larger contexts. In

Although this straightforward scholarly letter reports full familiarity with the substance of the Thomas theorem and with the eponym, it makes clear that such knowledge did not derive from having read the original formulation. Thus, the letter provides a distinct case of the basic difference between the *primary and the derivative or serial diffusion of ideas,* a subject of distinct interest to those of us at work on patterns of diffusion in science and technology (Coleman, Katz & Menzel 1966; Merton 1989, 1995; Rogers [1962] 1995; Zuckerman 1989). When a first-hand quotation does not provide an exact citation to its source, it cannot, of course, make for independent recourse to the original source and later derivative or serial diffusion can only reproduce the quotation as mediated. This, in turn, raises the sociological as well as normative question, which is here only introduced rather than explored: how does it happen that, unlike SSS who was ready to cite the mediated source, many of the authors who were plainly quoting the theorem derivatively rather than directly have not done so? In short, is there a norm for citing mediated references and if not, why not?

Is There A Norm for Citing Mediated References?

SSS's announced intention of citing his mediating source of the Thomas theorem may be defined by some peers as supererogatory, for there are evidently no well-established norms governing such citation behavior. At least, not if one may judge from the comparative rarity of citations to encyclopedias and other reference works in scholarly publications. Or more specifically, if one may judge from the frequency with which even mature scientists and scholars have written sociologists of science to ask whether or how to ascribe "proper credit to an author for drawing one's attention to a valuable reference."[14]

These normative questions about mediated references had not yet crystallized back in mid-century when the sometime sociologist of science Norman Kaplan (1965) introduced the generic problem of "the norms of citation behavior." Nor have these questions about *mediated references* yet been examined by the founder of the *Science Citation Index* and of citation analysis, Eugene Garfield (1983, 1995), in his longtime and still ongoing examination of how the norms and practices of citation are acquired. (See also Cronin 1984.) However, we now see how this detail in the collegial letter from SSS tacitly directs us to a generic, difficult, and important sociological problem: the dynamics of the emergence of a new social norm; (on the generic problem, see Coleman 1990). Just as another detail tacitly directs us to the institutionalized norm of socially organized skepticism.

---

this way, a book of quotations can extend an open invitation to the further reading or rereading of the original texts, beyond the quotations themselves" (Sills & Merton 1991: xvi).

[14]For example, the query quoted in the text which was addressed in the first instance to Eugene Garfield and derivatively to me by the author of "Thoughts on Eponyms," Howard B. Burchell, M.D. (1985).

SOCIALLY ORGANIZED SKEPTICISM

Having long ago identified this technical and moral norm embedded in the culture and the social structure of science (Merton [1942] 1973:267-78;339-40), I can only applaud SSS's manifest adherence to it. The term and concept, *socially organized skepticism,* refers to institutionalized arrangements for the critical scrutiny of knowledge claims in science and learning that operate without depending on the happenstance skeptical bent of this or that individual. The process of socialization in the culture of science joins with such social arrangements as published and unpublished "peer review" that serve as agencies of social control which see to it, among other things, that authors generally abide by the norm of indicating their predecessors and sources. That norm has many cognitive and social functions; Garfield (1983, 1995) lists 15 of them. Among the manifest cognitive functions are those of enabling scientists to consult prior sources to see whether they have been correctly utilized and whether they also provide pertinent information not included in the mediating source. A manifest social function is to pay homage to pioneers and other predecessors, along with its largely latent and correlative function of thus helping to maintain the reward system of science which, like all institutional reward systems, initiates or reinforces incentives for role performance.

Exemplifying socialization in this normative practice of organized skepticism, this committed scholar reports "having been well trained . . . in scholarly skepticism of sources," and goes on to use the secondary work of *Social Science Quotations* as a means of getting to the original source of the Thomas theorem that can then be examined at first hand. This soon led SSS to the discovery that the book in which the theorem first appeared was actually written by William I. Thomas *and* Dorothy Swaine Thomas and that, in turn, led (as we have seen that, in effect, it later led R.S. Smith) to the conclusion that ascribing the theorem to W.I. Thomas alone amounts to "institutionalized sexism."

This suggests the hypothesis that such a conclusion is especially apt to be drawn in a time of socially and culturally induced sensitivity to all manner of discriminatory and exploitative -isms. For, as the sociology of science has noted from its earliest days, "the question of the relative importance of intrinsic and external factors in the determination of the foci of scientific interest has long been debated" (Merton [1938] 1970:199) but there is no question that social and cultural contexts do variously influence problem-choice and hypothesis-choice (Zuckerman 1978, 1994). In that process, certain contextually influenced hypotheses rather than others soon leap to mind as plausible. In this case, it appears, the context-laden hypothesis that the failure to ascribe the Thomas theorem to both Thomases must be an expression of sexism. But of course that hypothesis raises the methodological question whether joint authorship of the book in which the theorem first appeared is enough to conclude that the theorem was itself a joint product. In self-exemplifying style, the cognitive norm of socially organized skepticism thus requires us in turn to examine that conclusion by probing other scholarly sources and personal archives, as we shall be doing in due course.

Meanwhile, one can see from the SSS letter how adherence to the norm of organized skepticism can lead to gratifying experiences that presumably

reinforce continuing adherence to it. Not least, the special pleasure that comes from erasing errors in received knowledge. For the detection of long-accepted error also has its social and personal functions. It not only contributes to the common wealth of science but results in system-induced heightened esteem by knowing colleagues. I too once found myself resonating to the kind of Eureka glow that evidently came to SSS upon discovering the unscholarly error that "not only [Sills] and Merton [but] indeed the entire U.S. (world?) scholarly community have made consistently." I still recall the scholarly thrill of first tentatively spotting and then actually demonstrating a centuries-long error in received scholarship. As it happens, this episode also involved the correct ascription of a memorable sociocultural aphorism (theorem?) to its actual author(s). The aphorism "If I have seen further, it is by standing on the shoulders of giants" — which had diffused widely in the literary as well as the scientific community, especially after it became known that Isaac Newton once made it his own — had long been ascribed to what I demonstrated was a phantom source in a work by the ancient poet Lucan rather than to its actual author, the twelfth-century Bernard of Chartres. Much along the same lines that SSS feared might be the case with the ascription of the Thomas theorem to W.I. Thomas alone in the then forthcoming Sills and Merton reference book, *Social Science Quotations,* the ubiquitous reference book Bartlett's *Familiar Quotations* had been perpetuating the error of an ancient source for the Newton aphorism — this, in no fewer than seven editions for almost a century. In the latter case, it became a scholar's comfort to find that even in this postmodern age of deconstruction, evidentiary truth can still prevail. For I can happily report that only thirty years or so after that ghastly error of a ghostly source had been demonstrated in *On the Shoulders of Giants* (Merton [1965,1985] 1993:246-60), it was finally corrected in the fine sixteenth edition of Bartlett's ([1882] 1992:281b) by its new scholarly editor, Justin Kaplan. And, in accord with the norm governing scholarly acknowledgment, the correction was made with due reference to that long drawn-out and digressive documenting work of mine.

But if the editor of Bartlett's *Familiar Quotations* saw the light and rectified what had been laboriously shown to be a fossilized error of faulty scholarship, not so, it appears, with the editors of *Social Science Quotations.* In unyielding style, they decided to continue ascribing the Thomas theorem to W.I. Thomas alone. This, despite the advance warning by their colleague-at-a-distance that to do so would not only perpetuate an almost universal error of quite unscholarly attribution but would amount to a "piece of 'institutionalized sexism.'" As we shall soon see in detail, that apparently intransigent decision was based on fortunate access to a personal archive.[15]

---

[15]This, however, is manifestly the place to anticipate the detailed documentary analysis a bit by replicating a memorandum on this decision dated "26 December 1989":

THE THOMAS THEOREM
Canonical Version

To: David
From: Bob

It seems to me that we should give readers or browsers the fruits of our unique documented knowledge that though the theorem appeared in the jointly written *The Child in*

Those recent observations by SSS and R.S. Smith on seeming inequities in the citation history of the Thomas theorem provide yet another specimen of a recurrent pattern in the domain of science and scholarship, to wit the pattern of multiple independent observations, discoveries, and inventions.

The Pattern of Multiple Independent Observations, Discoveries, and Inventions

As is well known, a longstanding theory of the development of science and technology is rooted in the theoretically strategic fact of the multiple and independent appearance of essentially the same scientific observation, discovery or invention; what, for brevity's sake, can be described as a "multiple" (Merton 1961, 1963). This pattern of independent multiples has been found to hold for all manner of cognitive contributions, ranging from the great through the intermediate and small to the trivial. An apparent example of the truly trivial multiple is provided in the triply independent observations by SSS (1988), Smith (1993), and myself (Merton 1976, [1948] 1982) that the Thomas theorem was being ascribed solely to W.I. Thomas, typically without even collateral reference to Dorothy Thomas. To situate the specific empirical case in theoretical terms, the widespread citation pattern of the Thomas *theorem* was identified at the outset of this retrospective article as an instance of the partial citation phenomenon — the delimited accrediting of a presumably joint contribution to a subset of the collaborators.[16] Whether this example of a multiple in the ongoing history of the Thomas theorem is as *theoretically* trivial as it is surely empirically trivial will become evident only upon examining the apparent counter-example of the typically full rather than partial citation of the important Ogburn--[Dorothy] Thomas (1922) paper on independent inventions. (On "the trivial and the important in sociology," see Merton [1961] 1973:59-62; 1987:16-19].)

America, it was nevertheless written solely by W.I. I won't burden you with another copy of the 14-page gloss on that fact when we were accused of 'institutionalized sexism' in having ascribed it wholly to W.I. in the Guidelines to Contributors.

I believe that the quotation should be ascribed to W.I. just as it is in the Guidelines. Plainly, a more extended arrow [our idiomatic term for explanatory notes appended to quotations] is needed to explain the lone ascription. It might read this way:

The Child in America 1928:572 — Although the 'Thomas theorem' appears in this book written jointly by W.I. Thomas and Dorothy Swaine Thomas, it is ascribed to him alone since Dorothy Thomas reported [insisted?] in a letter to one of the editors that she had done only the statistical portions of the book and that "the concept of 'defining the situation' was strictly W.I.'s."

Of course, this could be condensed — or extended!

[In the event, this explanatory note was neither condensed nor extended but substantially reproduced in *Social Science Quotations* .]

[16] As we shall see when we turn to the problem of explaining rather than merely identifying instances of the partial citation phenomenon, they often derive from the Matthew effect.

The Evocative SSS Letter in Retrospect and the RKM Letter in Prospect

SSS's succinct letter of 1988 along with the Smith oral publication of 1993 have thus provided much grist for the sociological mill as we focus on the normative question of proper ascription of the Thomas theorem. What SSS described as "an interesting footnote in the sociology of sociology" reflects various manifest and latent patterns of peer interaction that obtain within the institutional framework of science and scholarship. And as we shall now see, the extensive letter in response examines the *normative issues* raised by SSS, supplies previously unpublished archival evidence about the respective roles of W.I. and Dorothy S. Thomas in formulating the Thomas *theorem,* and addresses the theoretical and methodological problems involved in trying to explain this specimen of the *partial citation phenomenon.* It will come as no surprise, of course, to find that, like the SSS document, this one also exhibits patterns that have been identified in the sociology of science. And so, like the letter from SSS, herewith reproduced verbatim, it too will be subjected to ongoing commentary in discursive footnotes and bracketed text.

RKM TO SSS: A VERBATIM DOCUMENT ON THE THOMAS THEOREM AS AN INSTANCE OF THE PARTIAL CITATION PHENOMENON

SOCIAL SCIENCE QUOTATIONS

Who Said What, When, and Where

10 September 1988

Dear [SSS],

I want to add my thanks to David's for your willingness to put together a batch of quotations from X------ and Y-----. Knowing that I had been thinking of a much fuller annotation in SSQ for "The Thomas Theorem," David has asked me to respond to your "footnote in the sociology of sociology" concerning the proper attribution of the theorem. I'm glad to try my hand at that since it should help me move toward a proper annotation.

To begin with, you are surely not alone in having searched in vain for the prime source of the Thomas Theorem. As you say, I failed to give a specific reference when I first happened to refer to it as a theorem back in the 1930s and 40s. As a result, I've periodically received requests for the precise reference. I enclose such a fairly recent inquiry [as Exhibit 8], this one from Cynthia Epstein (who I know won't mind my including hers as a specimen document).

Now this is just the sort of information which David and I intend to have the scholarly apparatus of SSQ make instantly accessible. As you note, even our little

brochure, GUIDELINES TO CONTRIBUTORS, does provide "a proper citation" to the theorem. But then, having exercised a proper "scholarly skepticism of sources" by looking up the cited book source for yourself — this being the kind of behavior conforming to the norm of "organized skepticism" which I proposed as a major element in the social institution of science and scholarship back in the early 1940s — you go on to report your findings in these words:

Lo and behold, you [David Sills] are in error, an error which not only you but Merton [too] and indeed the entire U.S. (world?) scholarly community have made consistently. For the book was not written by W.I. Thomas, but by William I. and Dorothy Swaine Thomas. Nothing in the book indicates that he wrote some chapters and she others. They are joint authors. And, at least in the Social Science Quotations, this piece of (dare I call it?) 'institutionalized sexism' should not be perpetuated.

To make it clear that your observation registers a fairly newfound [general] sensitivity to the matter of sexism, you sign off as

"Yours in the new era,"

[It will be observed that in one respect, the otherwise clear-spoken Skeptical Social Scientist is here rather puzzling. Having noted in the first paragraph of the letter that the *Guidelines for Social Science Quotations* which Sills and I had sent along to potential contributors has "a proper citation" to the source of the Thomas theorem, that plainly assiduous scholar nevertheless goes on to declare that we, like "the entire U.S. (world?) community" have consistently erred in ascribing *the book* to W.I. Thomas alone. My response to this puzzle took this form:]

Now, you are altogether right of course in observing that the book in which the Thomas theorem first appears — THE CHILD IN AMERICA: BEHAVIOR PROBLEMS AND PROGRAMS (New York: Knopf, 1928) — was "not written by W.I. Thomas, but by William I. and Dorothy Swaine Thomas." Indeed, you will find that our SSQ brochure indicates as much in its list of sources of the specimen quotations on page 19. (It refers to W.I. and Dorothy S. Thomas, following the format of the INTERNATIONAL ENCYCLOPEDIA OF THE SOCIAL SCIENCES. But I allowed friendship to taint scholarly precision. All his friends referred to him as "W.I.", never as "William I." and W.I. clearly preferred that usage [as, we shall see, did his collaborator Dorothy]. But we should have cited him as he appears on the title page, not in this misplaced friendly but unscholarly fashion. As the editor of the IESS in which SOCIAL SCIENCE QUOTATIONS will appear as Volume 19, David may overrule me on this [as in the event, he did] since I note that the biographical entry in Volume 16 lists the book as having been written by W.I. Thomas and Dorothy S. Thomas. And I recall that the authoritative volume, edited by Ed Volkart for the Social Science Research Council back in the early 1950s, referred on the title page to the "Contributions of W.I. Thomas to Theory and Social Research," not to William I. Perhaps too many old friends of W.I. have improperly subordinated scholarship to friendship. But this is scarcely the matter central to our discussion. We are in thorough agreement, then, that the book was written by the two Thomases.

And you are also entirely correct in reporting that "Nothing in the book indicates that he wrote some chapters and she others. They are joint authors." There is nary a word in the book stating who thought or wrote what.

All this leads you to conclude that our ascribing the theorem to W.I. alone, rather than to both W.I. and Dorothy, is plainly a case of sexism which may become further institutionalized by the medium of SOCIAL SCIENCE QUOTATIONS. And you remind us that we have entered a new era where old-style sexism no longer goes.

Still, I am confident that as a scholar given to the careful sifting of fact from fable, you are ready to reconsider your initial interpretations in light of new evidence. In this case, your double indictment of David and me as guilty of scholarly error and of sexist bias to boot. With that in mind, I shall submit several[17] exhibits in the hope that you will want to withdraw both accusations (not that any of us are wholly free from error or immune from the contagion of inadvertent or of institutionalized bias).[18] For we both surely agree with those wise words of the 12th-century William of Malmesbury (much admired by my mentor, the dean of historians of science, George Sarton [1931, I:255] who wrote of him [that] "He was the best chronicler of his time; the first one after Bede who tried not only to chronicle events but to explain their causal relations"): "Throw out such dubious stuff and gird ourselves for a factual narrative." (Alas, my ancient notes fail to note the work in which this quotation [from Malmesbury] is to be found.)

Exhibit 1 *[The Thomas Theorem and the Matthew Effect]*

The first exhibit bearing on the case of the Thomas Theorem is drawn from my piece, "The Matthew Effect in Science," (SCIENCE, 5 January 1968, vol. 159, 55-63, as this was slightly amplified when reprinted in my collection of papers, THE SOCIOLOGY OF SCIENCE (University of Chicago Press 1973). By way of orientation, I should report that what I described as 'the Matthew effect' (after Matthew 13:12 and 25:29)[19] consists in the accruing of greater recognition by peers for particular scientific or scholarly contributions to scholars of great repute and the withholding of such recognition from [their collaborating] scholars who have not yet made their mark. Here it is being suggested that the Matthew effect might have operated in the very case which is of central interest to us at the moment. Thus:

> The problem of achieving a public identity in science may be deepened by the great increase in the number of papers with several authors in which the role of novice collaborators becomes obscured by the brilliance that surrounds their illustrious coauthors. Even when there are only two collaborators, the same obscurant effect may occur for the junior who exhibits several "inferiorities" of status. The role ascribed to a doubly or trebly stigmatized coauthor may be diminished almost to the vanishing point so that, even in cases of later substantial achievements, there is little recognition of that role in the early work. Thus, to take a case close to home, W. I. Thomas has often been described as the sole author of the scholarly book *The Child in America*, although its title page unmistakably declares that it was written by both William I. Thomas and Dorothy Swaine Thomas. It may help interpret this recurrent misperception to consider the status of the collaborators at the time the book was published in 1928. W. I.

---

[17]As the reader will soon notice, these documentary exhibits turned out to be a good many more than merely "several."

[18]SSS had no way of knowing that I would find this charge of potential "*institutionalized* sexism" particularly distressing. For it was back in the early days of the civil rights movement — a decade or so before the Civil Rights Acts of 1957 and 1960 — that I had attempted to identify and to analyze the social phenomenon of "*institutionalized* discrimination" as distinct from acts of discrimination by *individuals* (Merton 1948a: 120, 101). That had seemed to me a fundamental sociological distinction then as it seems to me still.

[19]By way of reminder: "Unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath."

Thomas, then 65, was president of the American Sociological Society in belated acknowledgement of his longstanding rank as dean of American sociologists, while Dorothy Swaine Thomas (not to become his wife until seven years later) was subject to the double jeopardy of being a woman of sociological science and still in her twenties. Although she went on to a distinguished scientific career (incidentally, being elected to the presidency of the American Sociological Society in 1952), the early book is still being ascribed solely to her illustrious collaborator even by ordinarily meticulous scholars.[19]

19. See the ascriptions of the book, for example, in Alfred Schutz. *Collected Papers*, 2 vols., edited and with an introduction by Maurice Natanson (The Hague: Martinus Nijhoff, 1962), 1:348, n71; Peter McHugh, *Defining the Situation* (Indianapolis: Bobbs-Merill Co., 1968), p. 7.

Exhibit 2 *[Why I do not take the Alfred Schutz ascription of the book to W.I. alone as ipso facto evidence of a sexist bias]*

The preceding quoted footnote indicates that even so meticulous a scholar as Alfred Schutz (1962:I, 348, n. 71) has managed to ascribe the salient *book* solely to W.I. To obviate any need for you to search out the passage in which he does so and thanks to our home photocopier, I Canonize it here (as I should perhaps have done in the published note):

> The "definition of the situation" refers to the so-called "Thomas theorem" well known to sociologists: "If men define situations as real, they are real in their consequences." [71]
>
> 71 It was first developed by William Isaac Thomas, in his book, *The Child in America: Behavior Problems and Programs*, New York, 1928, p. 572. See also W. I. Thomas, *Social Behavior and Personality*, edited by E. K. Volkart, Social Science Research Council, New York, 1951, pp. 14 and 80ff.; the term "Thomas Theorem" was coined by Robert K. Merton, *Social Theory and Social Structure*, Glencoe, 1949, 179.

You will notice at once the emphatic though probably unintended way in which, the to-me admirable scholar, Alfred Schutz, casts Dorothy Thomas into limbo by ascribing not merely the Thomas Theorem but the book to W.I. alone. Indeed, the fateful masculine pronoun in the phrase "his book" would seem to lend grist for your mill of sexism at work in the case of the Thomases. Nevertheless, as you see from my reference to Schutz's error in attributing the book wholly to W.I., I do not impute a sexist bias that made for the error. The reason? Our now fairly extensive studies of the Matthew effect in collaborations among scholars of greatly differing standing in the field have found this to operate quite regularly among male collaborators and among female collaborators as well. (Which is not to say, of course, that sex or gender does not affect the probability of achieving recognition for one's scientific or scholarly work; the intensive studies of men and women scientists by Harriet Zuckerman and Jonathan Cole over the years being the

basis for that statement.)[20] In view of all this, I adopted the clinical hypothesis that Schutz's ascription of the book to W.I. alone was probably an instance of the Matthew effect rather than a sexist predilection.

Exhibit 3 [*Wherein I reiteratively ascribe the theorem to W.I. alone and absolve the theorem of total subjectivism*]

This exhibit has me [once again] emphatically ascribing the theorem, *not* the book, solely to W.I., even though it appeared in a co-authored book. This I have done regularly since its logical character and, I confess, its assessed importance first led me to describe it as "The Thomas Theorem". (This was not, as Alfred Schutz understandably cites it in his foregoing footnote, in 1949 but back in 1938. But then, there was not the least reason for this colleague at-a-distance to know that the term had appeared earlier in my article, "Science and the Social Order" [Merton 1938:331-332]).

You will note that the following passage, again photo-copied to save you confirming search, connects the theorem to other formulations but singles it out as the most succinct and memorable formulation of much the same sociological idea. But I must not digress into questions of sociological theorizing. Here, then, is a pertinent fragment [drawn] from a piece of mine titled "Social Knowledge and Public Policy," first published in 1975 and reprinted in Merton [1976]; the fragment appears on pages 174-175, 177:

> **The Hazards of Subjectivism**
>
> The idea of the subjective component in human action has a long history in sociology and had an even longer history before we sociologists arrived on the historical scene. It is an idea, moreover, that has been formulated in various traditions of sociological thought: in the notion of *Verstehen* (roughly: intuitive understanding) advanced by Max Weber (and many others influenced by him), Robert MacIver's "dynamic assessment," Florian Znaniecki's "humanistic coefficient," Talcott Parsons's "voluntaristic theory of action," and Schutz's "phenomenological perspective." The idea was succinctly formulated by W. I. Thomas in what is probably the single most consequential sentence ever put in print by an American sociologist:
>
> If men define situations as real, they are real in their consequences.[20]
>
> Now, it is one thing to maintain, with Weber, Thomas, and the other giants of sociology, that to understand human action requires us to attend systematically to its subjective component: what people perceive, feel, believe, and want. But it is quite another thing to exaggerate this sound idea by maintaining that action is *nothing but* subiective. That extravagance leads to sociological Berkeleyanism (the allusion being, of course, to the English champion of philosophical idealism, not to an American geographic or academic place). Such total subjectivism conceives of social reality as consisting *only* in social definitions, perceptions, labels, beliefs, assumptions, or ideas, as expressed, for example, in full generality by the criminological theorist, Richard Quinney, when he writes that "We have no reason to believe in the objective existence of anything."[21] A basic idea is distorted into error and a great injustice is visited upon W. I. Thomas whenever his theorem is thus exaggerated.
>
> Exaggeration of a seminal truth produces its own brand of error. Total subjectivism, which maintains that only social definitions of the

[20]Cole (1979), Cole and Zuckerman (1984), [Zuckerman, Cole and Bruer (1991), Zuckerman and Cole (1994)].

> situation (or other subjective equivalents) determine the character of human action and its consequences, in effect manages to transform the Thomas Theorem into this fallacious maxim:
>
> > If men do *not* define situations as real, they are not real in their consequences.
>
> . . .To correct the imbalance that comes with total subjectivism and to restore the objective components of social situations to their indispensable place, we plainly need this counterpart to the Thomas Theorem:
>
> > And if men do *not* define real situations as real, they are nevertheless real in their consequences.

You evidently consider this ascription of the theorem wholly to W.I. (with an allusion to Dorothy Thomas only in a footnoted citation to the book)[21] as *prima facie* evidence of "sexism." In this case, however, presumably an expression of personal rather than "institutionalized sexism." But I urge you to withhold reiteration of this harsh impeachment until you have examined Exhibits yet to come.

Exhibits 4 and 5 *[Which bear renewed witness to an longtime effort to deter the ascribing of sexism to scholars who attribute the theorem solely to W.I.]*

A few prefatory remarks before I introduce these almost identical exhibits.[22] Nearly six decades ago, I elected to focus on social structures, social patterns, and social processes in the domain of science and scholarship — not a subject of immense scholarly interest back then. And, as you surely have no cause to know, some thirty years ago I first tried to identify the institutionally distinctive reward system of that domain since it seemed to me to provide part of the dynamics and patterning of scientific work. That was set out in some detail in a longish paper titled "Priorities in Scientific Discovery"(1957). Part IV of THE SOCIOLOGY OF SCIENCE (1973), designated as "The Reward System of Science," collects a variety of my articles focussed on the workings of that system. Among other things, that 1957 piece proposed the strongly stated hypothesis that contradictions

---

[21] In retrospect, I note that SSS and, for that matter, R. S. Smith in his 1993 paper, along with countless others before and after them might have obviated their continuing search for the provenance of the theorem had they happened upon Volkart's reprinting of the concluding chapter back in 1951 (Thomas 1951) or even this citation to *The Child in America* in 1975. But as George J. Stigler 's (1961) seminal paper, "The Economics of Information" led us to see both in principle and in fact, even in our new age of advanced information-technology, achieving retrieval of sought-for information can exact prohibitive costs in terms of time, energy, or money.

[22] This archival essay-letter, which was plainly not intended for publication, had adopted "the non-linear, advancing-by-doubling-back Shandean Method of composition" inaugurated by Laurence Sterne in his immortal eighteenth-century *Tristram Shandy* and hesitantly adapted, just two centuries later, in my own "Shandean Postscript," *On the Shoulders of Giants* (which, it may be remembered, turned up in the introductory pages of this paper as I was empathically resonating to the manifest pleasure expressed by SSS upon discovering what was taken to be the universal error of attributing the Thomas theorem to W.I.). Sterne's *Tristram Shandy* had drawn upon the technique of "stream of consciousness" long before William James had formulated that apt metaphoric concept in his monumental *Principles of Psychology* ([1890] 1950, I: 239) and longer still, before James Joyce and, to a degree, Virginia Woolf had put that technique to work in their novels.

between the reward system and the normative system of science made for such social pathologies as the occasional felonies of plagiarism and the creation of fraudulent data. (I stray from the subject in hand to mention this only because of the intense current interest, both scholarly and popular, in such pathological phenomena in science. Back in the 1950s, that effort to identify the varieties of misbehavior in science and to theorize about their systemic sources apparently seemed like little more than sociological muckraking.)[23]

Back to our immediate subject, I also report that Part V of the same volume, THE SOCIOLOGY OF SCIENCE, is given over to an examination of "The Processes of Evaluation in Science." As one would suppose, it includes the article on the Matthew effect and other pieces devoted to theoretically derived and empirically investigated patterns of [peer] recognition in science.

I report these longstanding interests, in supremely egocentric style, for a reason. These ¶s are designed to provide a context for the pair of exhibits, 4 and 5. From that context it can be inferred that I've had more than a friendly and ad hoc interest in the peer perceptions of my old, much admired, friends, W.I. and Dorothy Thomas. And so, as [the preceding exhibit and] those [following] exhibits testify, I have gone into print, from time to time, to get the story straight about the source of the Thomas Theorem.

The essential step, it seemed to me in light of certain patterns that occasionally emerge in the world of scholarship, was to provide a kind of anticipatory prophylaxis,[24] in print, directed against easy misinterpretations and harsh indictments of scholars who attributed the theorem to W.I. rather than to both W.I. and Dorothy. Perhaps getting the facts in print would preclude such indictments based on skimpy *prima facie* evidence. Thus:

Exhibit 4 *[Evidence that my own attribution of the Thomas Theorem to W.I. alone was neither 'sexist' nor a case of the Matthew effect]*

> 20. What we may call the Thomas Theorem appears just once in the corpus of W. I. Thomas's writings: on page 572 of the book he wrote with Dorothy Swaine (Thomas) Thomas entitled *The Child in America* (New York: Knopf, 1928). I ascribe the theorem to W. I. Thomas alone rather than to the Thomases jointly not because of his gender or great seniority but only because Dorothy Thomas has confirmed for me what many have supposed: that the sentence and the paragraph in which it is encased were written by him. There is thus nothing in this attribution that smacks of "the Matthew Effect," which in cases of collaboration between scholars of decidedly unequal reputation has us ascribe all credit to the prominent scholar and little or none to the other collaborator(s). On the Matthew effect. see Merton, *op. cit.*, 1973, Chapter 20.

(This wouldbe prophylactic footnote is quoted from the second 'edition' of the previously cited paper, "Social Knowledge and Social Policy," as it appears on page 175 of Merton, [1975] 1976.)

As you see, this note anticipates misinterpretations of my having attributed the theorem to W.I. alone. Having been immersed all those many years in the sociology of science, I had inevitably become aware of the frequent pattern of swift mis-imputations

---

[23] Early scholarly studies of the reward system of science include Gaston (1971), Hagstrom (1974), and Zuckerman (1977); for an analytical overview, see Zuckerman (1988:520-26). Drawing extensively upon sociological sources, the popular volume by the science journalists, Broad and Wade (1983), soon catalyzed the attention paid by the mass media to deviant behavior in science.

[24] On re-reading, I find the term "*anticipatory*" rather redundant.

of ideological or other bias. Hence, my evidently unavailing effort to counter in advance the charge that my attribution simply expressed (witting or unwitting) "sexism" or, in a parallel construction, inverted "ageism." Or that the attribution was simply another instance of the Matthew effect.[25]

I should pause to report that I erred in writing that "the Thomas Theorem appears just once in the corpus of W.I. Thomas's writings." The leading authority on that corpus, Edmund H. Volkart, wrote me recently to say that he had independently made the same erroneous claim. (A comparatively rare specimen of independent simultaneous error in contrast to the many known cases of independent simultaneous discovery.) Ed Volkart had learned from our longtime mutual friend, Eleanor Isbell — for so many years, the indispensable stalwart of the Social Science Research Council — that W.I. did put the theorem into print once again. This, in his essay, "The Relation of Research to the Social Process," which appeared in a symposium by The Brookings Institution in 1931 under the title, ESSAYS ON RESEARCH IN THE SOCIAL SCIENCES.[26]

Upon examining that essay by W.I., I see that he largely reiterates the original ¶ from THE CHILD IN AMERICA in which the theorem appears, clearly making no effort to elucidate or develop it further. Still, the fact, and it is truly a fact whatever the radical subjectivists might say, that the crucial sentence [re-] appears in an essay by W.I. [alone] would seem to provide publicly accessible evidence that he knew it to be his and his alone.

Speaking of reiterations in print, I would have you turn the page to Exhibit 5. This, as you will observe at once, is another wouldbe prophylactic note that largely repeats the one published a half-dozen years before. It appears in a revised introduction to "The Self-Fulfilling Prophecy," as it appears on page 248 of Merton [1948] 1982.

Exhibit 5 *[Which reports but does not document the fact that Dorothy Thomas herself had confirmed that she had nothing to do with the concept of "defining the situation"]*

> **What we have been describing as the Thomas Theorem appears on page 572 of the book he wrote with Dorothy Swaine Thomas in 1928: *The Child in America* (New York: Knopf). I ascribe the theorem to W.I. Thomas alone rather than to the Thomases jointly because Dorothy, who became Dorothy Thomas Thomas when they were married eight years after that book appeared, confirmed that the consequential sentence and the paragraph in which it was encased were written by him. Thus, nothing in this attribution smacks of "the Matthew Effect," which operates in cases of collaboration between scholars of decidedly unequal reputation to ascribe all credit to the eminent scholar and little or none to the collaborator(s)—supplementary note, 1982.

---

[25] In retrospect, I note here the pronounced but still undeveloped effort to avoid conflating the "partial ascription *phenomenon*" and possible hypothetical *explanations* of the phenomenon in terms of sexism, ageism, or, more generally, the Matthew effect. I trust that this will be clarified in the final section of this paper which examines the sociological and methodological import of both the phenomenon and its explanation.

[26] As we shall see, Thomas emphatically quoted the theorem a third and last time in Blumer (1939:85) but did not elect to apply or develop it further. That introduces a puzzle for future exploration. What are we to make of W.I.'s own sparse attention to the specific theorem in his ample writings on the governing concept of "definition of the situation"? The very question evokes Sherlock Holmes's observation on the methodological significance of the absent: "the curious incident of the dog in the night-time; the dog did not bark." (For the imperishable original, see Arthur Conan Doyle, [1894] 1953: 347; in perhaps more immediately relevant context, the methodological fragment reproduced in *Social Science Quotations* [Sills & Merton 1991:52]).

The reason for the largely reiterated footnote is by now surely clear to you as it had become to me. It was plain that a then-and-there explanation was needed whenever one attributed the theorem to W.I. alone. Else, some were bound to conclude that Dorothy was being robbed of this part of her ample intellectual contributions. This variant of the note no longer alludes to possible imputations of gender or age bias. Instead, having set forth the grounds for the attribution, it goes on to inform, or remind, readers that the collaborators had married — on 7 February, 1935 when he was 71 and she, 34 —and that Dorothy had been a Thomas before she married another Thomas.

That allusion to the post-marital "Dorothy Thomas Thomas" must seem altogether irrelevant to the subject in hand.[27] I agree. It is there only as an echo of Dorothy's letter to me in which she describes her role in the writing of "The Children [sic] in America" and reports that "The concept of 'defining the situation' was strictly W.I.'s." But readers of those prophylactic footnotes could not surmise all that. Would that I had been able to publish in full Dorothy's charming letter to me and my reply to her. Then the story would have been writ large and plain.

But if I could not intrude that correspondence in my latter-day articles referring to the Thomas Theorem, I can surely include it here.[28]

You have only to turn to the next page of this lengthening response to your own letter to David to find Dorothy's letter as Exhibit 6 and then turn to the page[s] following for Exhibit 7.

---

[27] R.S. Smith (1993: 4-5) found that three of the 40 textbooks referring to the Thomas theorem "imply she [DST] was married to W.I. Thomas" at the time it appeared in print. He notes that this was mistaken and goes on to observe: ". . . even if they were married, this begs the question of why include such information? According to *The Chicago Manual of Style* (1982), it is not scholarly practice to identify the marital status of coauthors. It is assumed that each author is a contributor in his or her own right and so making such a point is irrelevant or detracts from the purpose of citation. As Reinharz implies in the title of her article on Dorothy Swaine Thomas ("Wasn't She the Woman Married to William I.?") citing the work in this way reinforces the patriarchal practice of subsuming a wife's work under her husband's authorship." Thus led to the Reinharz article with its nicely ironic subtitle, one arrives at it only to find next to nothing there about either the personal life or the considerable worklife of DST. As this archival letter to the Skeptical Social Scientist makes plain in virtually anticipatory style, my own echoic allusion to their marriage no more exemplifies "the patriarchal practice" of subsuming "under her husband's authorship" Dorothy's virtual lifetime of pioneering work in demography (D.S. Thomas 1938, 1941; Kuznets & Thomas 1957-64) and her pioneering studies of the Japanese detention camps during World War II (D.S. Thomas & Nishimoto 1946; D.S. Thomas, Kikuchi & Sakoda 1952) than it exemplifies a tacitly "matriarchal" practice of subsuming under his wife's authorship W.I.'s virtual lifetime of work on his concept of "the definition of the situation." It is only that even meticulous scholars of Thomasiana, such as Janowitz (in Thomas 1966:xvii) and Tate (1974), unaware that DST had been "an ardent Lucy Stoner" — an American colloquialism for a married woman who insists on retaining her birthname — have assumed that Dorothy Swaine became Dorothy Swaine Thomas only after marriage.

[28] Though persuaded then as now that discursive footnotes providing correlatives and contexts serve a useful function, I hadn't the temerity to include Dorothy's letter as a rather extended footnote in papers that, unlike this one, were not focussed on the matter of attributions of the theorem to W.I. being taken as a (witting or unwitting) expression of sexism.

Exhibit 6 *[In which Dorothy Swaine Thomas supplies the archival smoking gun]*

GEORGETOWN UNIVERSITY
WASHINGTON, D.C. 20007

CENTER FOR
POPULATION RESEARCH

September 10, 1973

Professor Robert K. Merton
Department of Sociology
Columbia University
New York City, New York 10027

Dear Bob,

Many thanks for your nice letter and the copy of your paper on the Matthew Effect in Science. I have always enjoyed reading the things both you and Harriet Zuckerman have written. There is just one point in this article that puzzles me. I assume the underlinings refer to some index. If so, you apparently have me under the Swaine. I assure you I was born a Thomas and then married a Thomas who was no relation. I was an ardent Lucy Stoner and also swore I would never change my maiden name which I didn't.

In regard to The Children in America W. I. Thomas employed me as an assistant since he had been told by the Rockefeller group to get himself a statistian. The statistical portions were mine and I am sending you under seperate cover Volkart's book which makes this clear. The concept of "defining the situation" was strictly W. I.'s. With cordial regards,

Sincerely,

Dorothy

Dorothy Swaine Thomas
Professorial Lecturer

Exhibit 7 *[Reply to the smoking-gun letter by DST]*

CENTER FOR ADVANCED STUDY IN THE BEHAVIORAL SCIENCES

202 Junipero Serra Boulevard • Stanford, California 94305 *Telephone (415) 321-2052*

28 September 1973

Dear Dorothy,

I treasure your sentence and, unless you say no, I shall quote it on every possible occasion as a lovely piece of sociological history. The sentence:

> "I was an ardent Lucy Stoner and also swore I would never change my maiden name which I didn't."

As for the underlining that puzzled you, that is pure and, in this case, meaningless chance. I happened to xerox the version of the paper which appears in the page proofs of my forthcoming collection of papers, THE SOCIOLOGY OF SCIENC All the proper names in the book had been underlined for purposes of indexing (and I assure you that Dorothy Swaine Thomas appears in the Index as such and not as a Swaine).

During the almost forty years (!) since I first met W. I. and you, I have retained a happy image of the two of you together, in your every joint aspect. Evidently, that was more than a casual imprint.

As for the possible Matthew effect involving W. I. and you, I had only this in mind. True, you were very much the junior assistant but W. I., in his generous way, saw to it that you were identified as co-author on the title page. It is therefore of some interest that repeatedly THE CHILD IN AMERICA is cited as being by W. I. alone. Of course, that has nothing at all to do with W. I.'s concept of "defining the situation." That was a basic sociological idea for many when I was a freshman first becoming excited by this oddly-shaped field known as sociology.

Yours ever,

Bob

Robert K. Merton

RKM:ja

Dorothy Swaine Thomas
Professorial Lecturer
Center for Population Research
Georgetown University
Washington, D.C. 20007

[But before concluding this long response to SSS, still another intervention is needed by way of context: For lengthy though it is, this now archival document nevertheless did not include an essential part of Dorothy's and my correspondence; namely, my note signaling the pertinent passage in the "Matthew Effect" paper which led to her altogether unexpected singling out of W.I.'s and her own distinctive roles in writing *The Child in America*. Here, then, is that note as the only-now interpolated

Exhibit 7A *[which, along with Exhibit 1, had evoked Dorothy's smoking-gun reply that was reproduced as Exhibit 6 in the 1988 letter to SSS]*

31 August 1973

Professor Dorothy S. Thomas
Center for Population Research
Georgetown University
Washington, D.C. 20007

Dear Dorothy:

I had hoped to see you at the meetings and that is why I hadn't written you before now. And how wise you were not to come: 98° in the shade all through the sessions.

I am flattered to have you ask for a photograph of me but I don't have one at hand of me back in the days when I was president. But I'll see if I can dig one up and will happily send it on to you.

In the meantime, I thought you might have some interest in the enclosed paper if only because of what I say about you, W.I., and <u>The Child in America,</u> *on pp. 446-7.* [Italics have been added to underscore the relevant sentence that evidently evoked the unsolicited smoking-gun reply].

Yours,
Bob

It is of prime evidentiary importance that neither this note nor the indicated passage in the accompanying copy of "The Matthew Effect" (which appears as Exhibit 1 in the letter to SSS) refers to the partial citation of *the Thomas theorem*. As can be seen, the passage refers only to the partial citation of *the book*. It then proceeds to account for this instance of the partial citation phenomenon in terms of the Matthew effect, which is said to be all the more probable since the less-known collaborator is "subject to the double jeopardy of being a woman of sociological science and still in her twenties."[29] Yet, it will be noticed that *on her own initiative*, Dorothy Thomas takes occasion to partition their distinctive contributions to the book to emphasize that as W.I.'s statistical assistant "The statistical portions were mine and . . . [t]he concept of 'defining the situation' was strictly W.I.'s." The concluding portion of this retrospective article will collate the many statements, both public and private, to the same effect by both Dorothy and W.I. But first the conclusion of the long letter of response to SSS back in 1988.]

And there you have the essential documentation. These seven Exhibits could surely be extended in my old and new Thomasian files.* But perhaps this documentation is enough to indicate that SOCIAL SCIENCE QUOTATIONS will not be perpetuating any "institutionalized sexism" by attributing the Thomas Theorem to W.I. Thomas.

Yours, truly, from an era not quite so new,

Bob
Robert K. Merton

---

[29] Readers may want to refresh their memory by glancing back at Exhibit 1.

* . . . I append a few other Exhibits which might have been incorporated in the letter proper were this being put together on a word-processor rather than an old-fashioned Selectric II typewriter.

Exhibit 8:	Cynthia [Fuchs Epstein's] note of 21 May, 1981
Exhibit 9:	R.K.M. reply to C.F.E., 26 May 1981. [C.F.E. managed to put the gist of this exchange into her book then in press: *Women in Law* (Basic Books, 1981), 362n., thus providing another still-rare precise citation to the original source of the theorem.
Exhibit 10:	Title page and antecedent page of my copy of the first edition of THE CHILD IN AMERICA bought at the grand Leary's bookstore in Philadelphia on the 23d of June 1939

Exhibit 8

Columbia University in the City of New York

CENTER FOR THE SOCIAL SCIENCES

21 May, 1981

Dear Bob:

Checking through hundreds of copy-editor's queries on my book I had occasion to check ST &SS for the Thomas Theorem. To my surprise I find no footnote to the original "If men define things as real..."

I am convinced that if you chose to present it without footnote then you were applying some rule of common knowledge that makes the footnote unnecessary - indeed, perhaps vulgar.

So - I ask, even implore - for a spelling out of the rule (as I lay choked in dusty tomes pursuing thankless searches for forgotten footnotes).

Yours,

Cynthia

Exhibit 9

Columbia University in the City of New York | New York, N.Y. 10027

UNIVERSITY PROFESSOR   Fayerweather Hall

26 May 1981

Dear Cynthia:

Having a specific page reference* for the Thomas Theorem will in truth make your book quite distinctive. As you may have noticed, there are precious few cases - in fact, I don't recall many - referring to the famous sentence which do accurately pinpoint its source. This gives the impression that ST&SS may indeed have served as a conduit over the past thirty years for the Thomas Theorem. At any rate, I enclose the page reference.

I should explain that I attribute it to W.I. Thomas, rather than to both Dorothy Thomas and him because I once asked Dorothy about it. She was emphatic that both idea and formulation were entirely W.I.'s. Incidentally, she once told me also that part of the reason -- perhaps, only a very small part -- for her marrying W.I. is that it allowed her (as a Lucy Stoner) to retain both her maiden name and to take on her husband's name as well.

I write this just hours before Harriet and I take off on our much-awaited holiday in Italy and London.

With much love,

Yours,

Bob

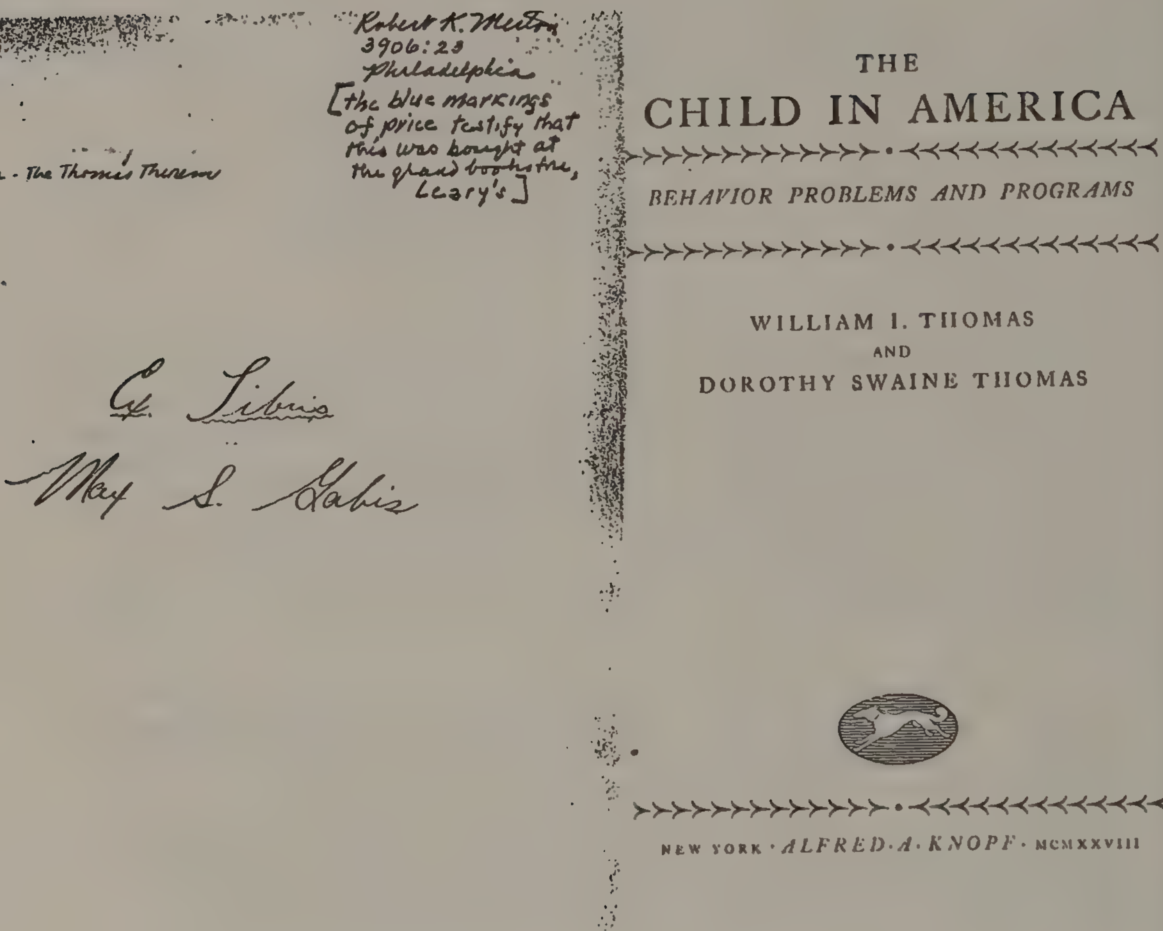
Robert K. Merton

Professor Cynthia Epstein
Center for Social Sciences
817 S.I.A.

* Encl. xerox of p. 175 in rkm, SOCIOLOGICAL AMBIVALENCE

Enc.

Exhibit 10


- The Thomas Theorem

Robert K. Merton
3906:28
Philadelphia
[the blue markings of price testify that this was bought at the grand bookstore, Leary's]

Ex Libris
May S. Gabis

THE
CHILD IN AMERICA
BEHAVIOR PROBLEMS AND PROGRAMS

WILLIAM I. THOMAS
AND
DOROTHY SWAINE THOMAS

NEW YORK · *ALFRED · A · KNOPF* · MCMXXVIII

Social Mechanisms Generating the Partial Citation Phenomenon

As indicated at the outset, this article is not primarily concerned with adjudicating claims to the origin of the Thomas theorem, and surely not with doing so on behalf of W.I. Thomas or Dorothy Swaine Thomas. Both simply took that origin for granted. We shall be examining further crucial evidence, from both a private archive and from publications, bearing on the recently disputed origin of the theorem, but our primary objective remains to understand the partial citation phenomenon and its place in the reward-system of science by analyzing this particular instance of that phenomenon as a strategic research site. The case of the Thomas theorem provides a strategic site if only because I happen to have first-hand and fine-grained archival information about the origin of the theorem, its early citations and subsequent citational history, the latter having been

usefully amplified by the Smith (1993) survey of the theorem in textbooks.[30] And though some of that archival material appears in the foregoing documentary letter to SSS, it leaves a variety of specific historical puzzles and generic sociological problems still to be specified and solved if we are to understand the specific case of partial citation and, through it, the general phenomenon. We begin with specific historical puzzles — how did it happen that the Thomas theorem was singled out for attention and how did it happen that its book source was not cited from the start? — and then attempt to account for the continuation of this pattern of citation in terms of a generic social mechanism in the cultural transmission of knowledge, namely, obliteration by incorporation.

Obliteration by Incorporation

To begin with, how did this one sentence come to be selected for incorporation in canonical knowledge from some 12,000± sentences comprising that book of 583 pages? After all, the Thomas theorem was neither the core subject nor the main theme of *The Child in America.* But if it was not central to the book, it *was* central — indeed it was a climactic formulation — for those social scientists in W.I. Thomas's invisible college[31] who for some time had been drawing upon his concept of *definition of the situation.* It crystallized a new phase of that evolving concept by adopting a pragmatic position to say much in little about the subjective component of action being truly consequential. And so members of that invisible college promptly fastened onto this new focus on consequences of definitions of the situation.

As was briefly noted in reviewing the early diffusion of the theorem, an integral member of that invisible college was the sociologist and social psychologist Kimball Young and, to the best of my knowledge, it was he who first isolated the sentence for conspicuous attention. This, it may be remembered, he did soon after publication of *The Child in America* by adopting it as an epigraph twice: the first time for Part Four of his textbook *Social Psychology* (1930: 397) and the next year for his chapter in an edited anthology, *Social Attitudes* (Young 1931:100). The first epigraph ascribes the quotation simply to "Thomas" *sans* given name or initials (just as the two other epigraphs on that same page are ascribed to [the tentatively inferred Ernest W.] Hobson and [the surely inferred

---

[30] It turns out that R.S. Smith has expanded his 1993 presentation to the Eastern Sociological Society and that this enlarged paper is scheduled for publication in *The American Sociologist* under the title, "Giving Credit Where Credit Is Due: Dorothy Swaine Thomas and the 'Thomas Theorem.'"

[31] In a felicitous stroke of terminological recoinage, Derek de Solla Price ([1963] 1986, passim; see Index) adapted Robert Boyle's seventeenth-century term "invisible college" to designate past and present informal collectives of closely interacting scientists limited to a size "that can be handled by inter-personal relationships." See also Diana Crane, *Invisible Colleges* (1972).

John] Wesley). The second book, which the dedicated Young[32] had dedicated to his master, W.I., ascribes the quotation more definitely to "W.I. Thomas." But as is commonly the case with epigraphs, neither of these gave a bibliographic reference to its source. Thus began the unpremeditated practice of quoting the Thomas sentence *sans* the full scholarly citation.

In examining its early diffusion, we also noted that the first descriptions of the sentence as "the Thomas theorem" likewise provided no bibliographical citation. And this for reasons as happenstantial as Young's having first quoted the sentence in epigraphs. The first such description of the quoted sentence as "W.I. Thomas's sociological theorem" (Merton 1938:333) was merely a passing allusion that clearly required no citation while the first detailed reference to the truncated "Thomas theorem" (Merton 1948) appeared in *The Antioch Review*, a journal for the common reader which did not look kindly on footnotes in general and on bibliographical citations in particular.[33] (The one citational footnote in this article was not to the theorem but referred to a publication that appeared while proofs were being corrected and was barely negotiated into being.) And when the article was reprinted, as it often was, the citation remained absent.

Thus it was owing to peripheral and surely unplanned circumstances that these early appearances of the theorem in print did not include the usual academic citations of its source. This chanced feature of its primary diffusion from members of the invisible college to the larger community of sociologists became serially reproduced in the course of secondary diffusion through textbooks and correlative writings. Thus there developed a special eponymous variant of the social mechanism known in the sociology of science as "obliteration by incorporation" or by the ultimate brevity, OBI, an acronym that stands for Obliteration of the source of ideas By Incorporation in currently canonical knowledge (Merton 1968: 27-28, 35-37; Garfield [1975] 1977).

This is described as a "special eponymous variant" inasmuch as OBI ordinarily involves obliteration of *both* the author(s) *and* the original source. This unplanned social mechanism consists in "users and consequently transmitters of the particular bit of knowledge [having] become so thoroughly familiar with its origins that they assume this to be true of their readers as well. Preferring not to insult their readers' knowledgeability, they no longer refer to the original.

---

32 At the time Young was seizing upon the theorem for his epigraphs, he also wanted to do a biographical piece on W.I. who emphatically refused to give him leave. Characteristically, he responded to Young's urgent request in this vigorous fashion: "I don't regard myself as important. I don't want to be noticed. I don't care whether a word appears about me in print, living or dead." Unhappily, Thomas's wish remains fulfilled: there is still no full-scale biography of this founding American sociologist. Manuscript letters by Kimball Young (30 April, 1930) and by W.I. Thomas (4 May, 1930). I am indebted to David L. Sills for making these letters available to me. Although Kimball Young's newly published oral memoir (1995) is at times fallible when unchecked by supporting documents, this deftly edited account provides much detail about the W.I. and the Dorothy Swaine Thomas sociocognitive network.

33 None of the other five articles in that June 1948 issue of *The Antioch Review* contained a citational footnote; there were five discursive footnotes all told. The next (September) issue consisted of 17 articles, with a total of six citational footnotes.

source" (Merton 1993:218-20) as is the case, for example, with such diverse knowledge-units as the theory of games, secondary analysis, revealed preference, opportunity costs, or latent structure analysis.

At the tail end of the OBI mechanism, many concepts-and-phrasings — such as charisma, stereotype, lifestyle, significant others, double-bind, and role-model among countless others — enter the vernacular, with rare if any awareness of their sources in the social sciences (Merton 1982:100-106). Obliteration by incorporation in the general culture set in long before for much-quoted dicta and concepts such as Francis Bacon's "knowledge is power" (1597), Joseph Glanvill's "climate of opinion" (1661), and John Adams's "government of laws and not of men" (1774). But plainly, in those cases where a scientific contribution has been eponymized, as with Boyle's Law, Halley's comet, Le Chatelier's principle, the Rorschach test, Gini coefficient, or Thomas theorem, only its original bibliographic source and not its author or authors becomes obliterated. I hazard the hypothesis that when the new idea is not eponymized, its source is more probably, more rapidly, and more extensively, deleted.

Thus, in his fine account of "the Chicago School of Sociology," Lester Kurtz (1984:34) in effect exemplifies both types of obliteration — of publication only and of author too — as he reports the differing fates of Thomas's uneponymized and eponymized concepts:

> The most persistent of [W.I.] Thomas's specific contributions is his emphasis on interaction and situations in the study of the subjective side of social life. His concept of the 'definition of the situation' has become one of those concepts so widely used in sociological analysis that it is often not explicitly attributed to Thomas. His situation analysis is a result of the influence of pragmatism, and much of his later work can be encapsulated in his phrase 'If men define situations as real, they are real in their consequences' (Janowitz 1966, p. xl). Much has been made of the formulation, called the 'Thomas theorem' by Merton (1968, pp. 475ff.), including efforts to link it with dramaturgical and ethnomethodological perspectives . . .[34]

To be necessarily obvious about it, when authors fail to provide a citation to the original source of a concept or quotation, this cannot directly lead their

---

[34] Along with providing apt examples of the two degrees of obliteration — complete obliteration of the source of the generic *concept* (definition of the situation) and inevitably only partial obliteration of the source of the *eponymous theorem* — this brief passage provides a singular array of well-identified patterns in the transmission of knowledge. Thus, the thoroughly knowledgeable scholar Kurtz elects to cite a secondary rather than primary source of the theorem, taking care to abide by an only slightly institutionalized norm of citation by citing that mediating source in precise detail. Degrees of obliteration in the transmission of knowledge are also exemplified as Kurtz draws upon Janowitz's observation that "much of this later work [by W.I. Thomas on what 'he called situational analysis'] was encapsulated in *his* [*n.b.*] phrase, 'If men define situations as real, they are real in their consequences,' as stated in *The Child in America*." (Janowitz 1966:xl). But then we find that authoritative scholar of the Thomas corpus, Morris Janowitz, not pausing to give a full citation indicating where the Thomas sentence is to be found in that expansive 583-page volume. And as a final irony in Kurtz's meticulous passage, we note that though he cites only a mediating source for the consequential sentence itself, he carefully cites the exact primary source of its eponym, "the Thomas theorem."

readers to that source. Such omissions make for the type of serial diffusion that merely reproduces the pattern of an uncited concept or quotation. Failure to cite a source need not, of course, result only from chance circumstances such as those at work in the early history of the Thomas theorem. It may also result from a conflict between diverse social and cognitive functions of citation. As the reward system of science evolved, paying peer respect by citing one's predecessors became essential to it, but plainly it would be highly inefficient were scholars required to specify the origins of every bit of incorporated knowledge every time it was put to use. As Kaplan (1965) and Garfield (1995) have in effect shown, the norms governing citation practice are still neither sufficiently detailed nor standardized to solve this problem of conflicting functions of citation.

At any rate, the Thomas theorem suffered a deficit rather than such excess of precise citation. This, owing to the chance of its first quotations by members of the Thomas invisible college having appeared in the form of epigraphs and eponyms that failed to signal its precise source. That omission was reproduced through later serial diffusion among authors who had themselves evidently not come upon the original source in *The Child in America*. Despite the later precise citations of the theorem by members of the Thomas invisible college (starting with the Volkart anthology of W.I. Thomas in 1951), that outcome of obliteration except for the eponym is, to judge from the Smith (1993) survey, still typical of textbooks introducing students to sociology. Obliteration by incorporation in the transmission of knowledge remains a largely intact process until subjected to exogenous historical or sociological examination.

But if these social mechanisms of OBI and serial diffusion account for the widespread failure to cite the book source of the theorem at all, they plainly do not account for this double case of the partial citation phenomenon which has W.I. Thomas being solely credited not only with the theorem but also with the book in which it appeared. Each of these types of partial citation derived from distinct though mutually reinforcing social mechanisms in the transmission of knowledge and so requires separate analysis.

Primary Group Ascription of the Theorem to W.I.

There was nothing problematical about the origin of the theorem among members of the Thomas sociocognitive network and particularly among those who at the time the theorem appeared or soon afterwards were in close touch with the central figures of that network, W.I. Thomas and Dorothy Swaine Thomas.[35] Members of the network who have entered this account as early

---

[35] Much tacit and explicit knowledge was exchanged in such networks then as now, Those networks were, of course, less numerous and less specialized back in the 1920s and 1930s when the membership of the American Sociological Society numbered some 1200, less than a tenth as many as the membership of the ASA in the 1970s and 1980s. (I am indebted to Valerie Pines of the ASA for this information.) One-time members of the Thomas networks are of course in increasingly short supply; my own relatively youthful engagement with the Thomases did not begin until W.I. came to Harvard as a visiting professor in the mid-1930s, the year I became a newmade instructor.

disseminators of the theorem — Kimball Young, Howard P. Becker, Willard Waller, and myself — had ample cause to regard the theorem simply as a memorable formulation of a new aspect of W.I.'s longstanding paradigmatic idea of definition of the situation. Which we know was also the case with Dorothy Thomas herself since, without any prompting on my part, she had volunteered as much in her smoking-gun letter (cf. Exhibit 6).

However, what was transparently obvious at the time to Thomas colleagues close-by and at-a-distance has evidently become problematic in this time of political correctitude — at least for those who define the sole ascription of the theorem to W.I. Thomas not as a matter of historical fact but simply as an expression of sexism. It may therefore be useful to collate a few more bits of evidence from both public print and private archive that bear upon this current definition of this particular situation as further prelude to an analysis of the widely diffused partial citation of the book and the theorem.

DST'S RETROSPECTIVE ON *The Child in America*

Some twenty years before Dorothy Swaine Thomas had crisply and emphatically described W.I.'s and her own distinctive contributions to the book in the smoking-gun letter, she had taken the occasion of her Presidential Address to the American Sociological Society to do so in much greater detail. An incomplete private archive yields an almost instant response to the Address that impatiently aimed to spread the word just a bit more quickly in a way typical of sociocognitive networks.

26 September 1952

Dear Dorothy:

Thinking back, I find that I'm glad to be a member of the Council. I am sure I would have seen nothing of you at the Atlantic City meetings if I hadn't been among those fortunate few who spent almost three whole days in your private suite. I've told Paul Lazarsfeld about your presidential address and if you have a spare copy of it, could you send it on for him to read so that he will not have to wait for publication?. . .

The response follows soon:

October 3, 1952

Dear Bob,

I too enjoyed the three days of Council meetings. Thanks for your kind remarks. I do not have an extra copy of the Presidential Address, but it will be published in the December issue of the Review. Under separate cover, I will send you, within a few days, a copy of a very good picture of W.I . . .

What, then, does Dorothy Thomas (1952: 665) have to say about our subject in that Address by the first woman to become president of the then ASS? Herewith, a few snippets from this autobiographical piece on the life of the mind (that still bears re-reading):

The framework of *The Child in America* was W.I. Thomas's famous situational approach . . . which defined the 'total situation' as always containing more or less of the subjective. . . It is always dangerous to try to reconstruct the separate contributions of

collaborators, but I am reasonably sure that the designation of subjective, documentary materials as the 'as-yet unmeasured' and the emphasis on 'transmuting' more and more factors 'into quantitative form' were mine and that the very positive evaluation of the behavior document *per se* was W.I. Thomas'. For when I joined the staff of the Child Development Institute at Teachers College in 1927, I was still somewhat distrustful of the subjective and the 'as-yet unmeasured' as materials for scientific investigations. I still preferred to work with the objective, defined in almost mechanistic terms and to count, measure, sample, fit curves, correlate, test for reliability, validity and the significance of quantitative differences, rather than to utilize descriptive materials or life histories, case records, and other types of personal documents. I hoped, indeed, that the series of observational studies of social behavior which I directed there and continued during the 1930's at the Yale Institute of Human Relations might yield 'data as objective as the best of those with which the statistical economists' were dealing. And although I gave verbal recognition to the value of case histories, diary records, and what I called 'merely descriptive' accounts of behavior as 'hypothesis-forming material for further studies' I made slight use of these materials, on the ground that they 'obviously [would] not yield data appropriate for statistical analysis' (D.S. Thomas 1952: 665 citing D.S. Thomas 1929: 19-30, passim).

This public avowal with its emphasis on Dorothy Swaine Thomas's commitment to 'objective' statistical analysis inevitably brings to mind her statement in the 1973 smoking-gun letter that "the statistical portions [of *The Child in America*] were mine." That specialized role was in effect reaffirmed by W.I. (who insisted on declaring his technical ignorance of statistics) when he concluded one of the only two papers in which he reproduces the theorem by stating that "Dorothy Swaine Thomas, of Yale University, is responsible for the items relating to statistical procedure in this article."

DST'S TACIT APPROVAL OF VOLKART'S 1951 ASCRIPTION TO W.I.

At almost the same time, Dorothy Thomas is providing behavioral testimony to the division of scientific labor in the book *and* to virtual ascription of the theorem to W.I. Four years after W.I.'s death in 1947, she is first among many who are troubled by the lack of access to his works "in the field of social behavior" that were out of print or had never been published (Thomas 1951:xi). She draws upon three colleagues, near and far, in her own sociocognitive network — Donald Young, Thorsten Sellin, and Herbert Blumer — to serve with her as a committee of the Social Science Research Council to oversee the collection of W.I.'s writings that was to be brought together by Edmund Volkart (Thomas 1951). Volkart of course elects to reprint the consequential final chapter of *The Child in America* (duly cited as written by the two Thomases) and observes in his introduction to it that "the importance of subjective experience to a science of behavior is still emphasized." That continuing emphasis on the subjective is manifestly being ascribed to W.I. and surely not to Dorothy Swaine Thomas for, as we know, she is just then reporting in her presidential address her (since relaxed) "distrust of the subjective" at the time the book was being

written.[36] Volkart goes on to note, as a matter of course, that "In this connection Thomas' [*n.b.*, not "the Thomases'"] discussion of the life history as a source of research material should prove especially useful to students of culture and personality." As a member of the committee supervising the volume, Dorothy Thomas evidently did not object to this reading of the chapter containing the paragraph on the value "of the highly subjective record . . . for behavior study" which culminates in the theorem. Once again, we are put in mind of her smoking-gun letter which resonates with that reading of the chapter when she writes that "the statistical portions were mine and I am sending you under seperate [sic] cover Volkart's book which makes this clear. The concept of 'defining the situation' was strictly W.I.'s."

WIT'S SMOKING-GUN ASCRIPTION OF THE THEOREM TO HIMSELF

W.I. evidently agreed. For long before, he had anticipated Dorothy Swaine Thomas's private smoking-gun letter in a public smoking-gun ascription to himself while addressing a panel of social scientists appraising Blumer's critique of *The Polish Peasant*; this, in the course of yet again stating his methodological case for "the behavior document, whether autobiography, case record, or psychoanalytic exploration."[37] In a matter-of-fact vein (here italicized for emphasis, no doubt needlessly), he remarks that "*I* quote what *I* said in this connection" and then proceeds to quote the passage from *The Child in America* we have come to know so well, the one that closes with the now symbolically historic sentence: "If men define situations as real, they are real in their consequences" (W.I. Thomas in Blumer 1939:85).

DOCUMENTARY CONFIRMATION OF THE DST AND WIT SMOKING-GUN ASCRIPTIONS

Even in the absence of these archival and published documents by both Thomases, intellectual historians would have little difficulty in ferreting out the

---

[36] I write "since relaxed distrust of the subjective" inasmuch as the presidential address concludes, with typically Thomasian candor: "On the behavioral side, I have not found it profitable to proceed as if all behavior must be or even can be 'transmuted' into quantitative terms. And whereas I still push the statistical aspect of all studies to the limit, I no longer relegate the subjective and the descriptive to secondary positions" (Thomas 1952: 669). It was that kind of candid public retrospective that elicited this fan letter drawn from my own private archive: "I salute our out-going President — this being said as a sociologist and referring therefore to the Chief Executive of the A.S.S rather than the U.S.A. The particular occasion for my drinking to your good health is the appearance, in print, of your salty, meaty and otherwise nourishing presidential address. It stands up on the printed page as it did before the collected audience. Most of us never try to make sense of the life of the mind we have led, and of the few who do try fewer still succeed. Yours was a complete success, all the more solid for being wholly unpretentious" (RKM to DST, 27 December 1952).

[37] This, it should be noted, is a virtual quotation from a comparable passage in *The Child in America* which reads: "The behavior document (case study, life-record, psychoanalytical confession. . ." (571)

highly probable authorship of the theorem. For neither before nor after publication of *The Child in America* did Dorothy Thomas make sustained use of the theorem or the more inclusive concept of definition of the situation whereas W.I. Thomas devoted much of his twentieth-century worklife to what he described as "situational analysis." In 1917 — Dorothy Swaine Thomas was then 17 and about to enter Barnard College — his influential paper on "The Persistence of Primary-group Norms in Present-day Society"[38] observes that "this defining of the situation is begun by the parents. . ." and a few years later, *The Unadjusted Girl* (Thomas 1923) is "mainly concerned with situational analysis and the definition of the situation" (Janowitz in Thomas 1966:xxvii). But rightly enough, it is his presidential address to the American Sociological Society, a year before publication of the for-us landmark book, *The Child in America,* that almost wholly anticipates the formulation of the theorem:

> A document prepared by one compensating for a feeling of inferiority or elaborating a delusion of persecution is certainly as far as possible from objective reality. On the other hand, this definition of the situation is from one standpoint quite as good as if it were true. It is a representation of the situation as appreciated by the subject, "as if" it were so, and this is for behavior study a most important phase of reality (Thomas 1927:7).

There it is: the essential idea, down to the detail of including the Hans Vaihinger ([1911] 1924) phrasing of "as if" as this appears in the canonical version of the theorem. This anticipatory version lacks only the pragmatic element of consequences made explicit and the felicitous formulation that made the theorem memorable.

In light of this cumulation of private and public evidence, it does not seem extravagant to conclude that Dorothy Thomas and W.I. Thomas were probably speaking truth about the origin of the theorem.

ZNANIECKI'S SUMMARY OF THE WIT AND FZ DIVISION OF SOCIOLOGICAL LABOR

This accumulation of evidence would also seem to bear upon the suggestion by Howard P. Becker that the theorem was "in content at least . . . probably Znaniecki's." Beyond this evidence is Znaniecki's statement to that panel of social scientists engaged in appraising Blumer's appraisal of *The Polish Peasant.* There he summarizes Thomas's and his own "previous results of comparative analysis and generalization," thus:

> Thomas had at the time already formulated several well known and original theories in social psychology and sociology, based upon an exceptionally great mass and variety of significant data carefully chosen from many different cultures; and in starting to collect materials concerning European peasants he meant to apply his theories to this new mass of data. I had published several works in general theory of culture and in epistemology which eventually proved to have some bearing, however abstract and indirect; the former

---

[38] Thus, in an article anatomizing textbooks in social pathology, C. Wright Mills (1943: 171) notes that "About the time W.I. Thomas stated the vocabulary of the situational approach, a social worker was finding it congenial and useful. In M.E. Richmond's influential *Social Diagnosis* (1917) we gain a clue as to why pathologists tend to slip past structure to focus on isolated situations. . ."

upon the data of Polish peasant culture, the latter on the method of handling them. (Blumer 1939:87,90)

Thus, typical of both, Thomas takes the occasion to focus anew on his concept of "defining the situation" in general and on his theorem in particular while Znaniecki has nary a word to say about concept or theorem. Instead, he straightforwardly distinguishes "Thomas's well known and original theories in *social psychology and sociology*" (which, of course, were centered on his idea of "definition of the situation) from his own "*theory of culture and epistemology*." In this way, Znaniecki is also being forthright for, like Dorothy Swaine Thomas, he too made no sustained use of the idea either before or after the great collaboration.[39]

So much, then, about how it was that members of the Thomas sociocognitive network ascribed the theorem solely to W.I. Given their first-hand knowledge of its origin, no alternative could possibly have occurred to them. But, as I have emphasized from the start, this article is primarily concerned with examining the Case of the Thomas Theorem in an effort to understand the generic partial-citation phenomenon, not with the specific matter of adjudicating its origin. And though early obliteration by incorporation and serial diffusion may explain the partial citation of the *theorem* by the authors of textbooks and others who had no direct access to this first-hand information, those social mechanisms cannot, of course, explain the practice of ascribing the *book* solely to W.I., since that was not a practice appearing often in the early diffusion of the theorem.

However, it may be remembered that documentation supplied in the archival letter to SSS briefly proposed that this specific case of partial citation of the book may have resulted from the Matthew effect operating as a generic social mechanism in the transmission of knowledge. That proposal warrants further theoretical and methodological scrutiny in light of the concern voiced by SSS that ascribing the theorem to W.I. was sexist and the renewed interpretation by Smith (1993) that this practice "can also be explained in terms of a structural issue — the genderization[40] of the discipline as part of the process of professionalization. By not citing Dorothy Swaine Thomas these authors help sustain a view of sociology as historically a male domain."

---

[39] In striking contrast to Thomas who, in his last, massive, work, *Primitive Behavior* (1937), draws upon the favored mode of situational analysis from its first page to its last. The very first sentence in the book announces that every aspect of culture "can best be approached in terms of *the definition of the situation*" while the index lists 28 pages which explicitly deploy that concept.

[40] The "genderization" of scientific disciplines has been described by Evelyn Fox Keller (1985) as involving the domination of the sciences by men scientists at the expense of recognition of women scientists. Keller (1991) has gone on to put her argument thus: "the exclusion of the feminine from science has pertained to a particular definition of science: science as incontrovertibly objective, universal, impersonal — and also masculine" (235).

The Partial Citation Phenomenon and The Matthew Effect

As the terms imply, *partial citation phenomenon* designates a fact while the *Matthew effect* designates a theory. There is no reasonable question that incomplete citations of authors occur but it remains to be shown whether and to what extent and under which conditions they are cases of the Matthew effect. This is not the place to attempt a further systematic explication of the effect but, as noted earlier, there has been a tendency to conflate the fact and the theory, the phenomenon and its proposed explanation.

In cases of collaboration between scholars of notably unequal reputation, the Matthew effect confers excessive credit on the better-known scholar(s) and little or none on the other collaborator(s). Thus, the biologists R.C. Lewontin and J.L. Hubby instructively describe the far larger numbers of citations accorded one of a pair of their joint papers as apparently

> a clearcut case of [the] 'Matthew Effect' . . . In 1966, Lewontin had been a professional for a dozen years and was well known among population geneticists, to whom the paper was addressed, while Hubby's career had been much shorter and was known chiefly to biochemical geneticists. As a result, population geneticists have consistently regarded Lewontin as the senior member of the team and given him undue credit for what was a completely collaborative work that would have been impossible for either of them alone. (Lewontin & Hubby 1985:16)

This report holds immediate interest for us here since in this case, as in countless others, the Matthew effect cannot be easily attributed to a difference of gender. After all, their fellow scientists know that both Lewontin and Hubby are males. The fact that partial or other forms of skewed citation most often refer to author-sets of the same gender is enough to raise the interesting theoretical question why, in the case of the Thomas theorem, failures to include Dorothy Thomas in citations of the theorem or the book have been promptly attributed to her being a woman and his being a man. Both W.I. and D.S. Thomas occupied complex status sets.[41] He was not merely another male sociologist nor she merely another female sociologist. They had many other differing statuses and distinctive social attributes. For example, the attribute of their comparative standing in the discipline and in the field of social psychology (in which Dorothy Swaine Thomas had not worked at all). At the time that one memorable sentence appeared in the coauthored *Child in America* and diffused as the Thomas theorem, W.I. was widely recognized as one of the most consequential of American sociologists. (A quick and, for historians of sociology, redundant indicator of his standing was provided by a study of the comparative amount of space devoted to founders of the discipline in historical textbooks that found him ranked first among living sociologists and sixth among the likes of Durkheim, Comte, Spencer, Ward, and Max Weber [Palmore (1962) 1971].)

---

[41] On the structural concept of status-set as the set of distinct socially defined positions occupied by individuals at a given time, see Merton (1968:434-38); for its application to the phenomena of "sex typing" and structural obstacles confronting women in American society, see Epstein (1970:86-101; 1988:101-28).

Thomas was esteemed in particular for his theoretical development and application of situational analysis. At the age of 64, just a year before publication of *The Child in America*, he had finally been elected President of the American Sociological Society (presumably because enough members had abandoned their prudish attitudes toward his alleged notorious adultery that had led to his dismissal from the University of Chicago ten years before). Along with his exalted place in the social stratification of American sociology and other statuses in his status set, such as his age, race, religion, ethnicity, university affiliation, W.I. Thomas also happened to be a man.

In contrast to W.I. Thomas's worldwide fame, Dorothy Swaine Thomas had yet to achieve her fame. And, as was noted a quarter-century ago (see Exhibit 1), she was also "subject to the double jeopardy of being a woman of sociological science and still in her twenties."

With such enormous differences in the extent of obviously age-related accomplishment and reputation between the two coauthors, why should we fasten onto the one status difference of gender to explain this case of partial citation? On what grounds should we assume that this one status determined both the amply warranted ascription of the theorem and the wholly mistaken ascription of the book to W.I. Thomas alone? To the extent that gender-, race-, age-, ethnic-, or other status-influenced ascriptions do enter into particular cases of the partial citation phenomenon that plainly involve the Matthew effect, they involve overdetermination (in the methodological, not the psychoanalytic, sense of having more determining factors than the minimum necessary to bring about the outcome). All this raises a series of theoretical and methodological questions.[42] How do we go about discovering whether and to what extent cognitively irrelevant statuses of authors and of peer ascribers make for the partial citation phenomenon? Do scientists tend to attribute the prime role in collaborative work by men and women to those of their own gender? If so, does this practice differ by gender? Does it obtain irrespective of the comparative standing of the collaborators generally and in their special fields of investigation? In short, there is still much to be done by way of systematic empirical investigation of the diverse workings of the Matthew effect in relation to such functionally irrelevant statuses.[43]

---

[42] These generic theoretical and methodological questions also hold specific personal interest. For, as Eugene Garfield (1994:13) has recently reported, my colleague Harriet Zuckerman and I have long been subjected to a pattern of mis-citation reminiscent of the Lewontin-Hubby experience with the skewed distribution of citations to their two joint papers. Papers written jointly by "Zuckerman and Merton" (1971, 1972) are often cited with the order of the authors reversed. This, of course, carries its own irony, since the author who Garfield notes "had identified, named, and harshly criticized" the Matthew effect thus becomes its dubious "beneficiary." The irony becomes all the greater in light of a statement inserted in a reprinting of the 1968 paper, "The Matthew Effect in Science": "It is now [1973] belatedly evident to me [RKM] that I drew upon the interview and other materials of the Zuckerman study to such an extent that, clearly, the paper should have appeared under joint authorship" (Merton [and Zuckerman] [1968] 1973:439).

[43] For inquiries into the workings of the Matthew effect, though without reference to this matter of gender- or other status-influenced attributions, see Stephen Cole (1970; 1992: chap. 6).

In their normative aspect, the partial citation phenomenon and Matthew effect make for injustices. They violate the basic norm of giving peer recognition of contributors to the common wealth of science and scholarship. And, as we have seen, such normative violations evoke deep-seated responses. To extend the normative problem in the case of collaboration among status unequals, the tendency toward this type of injustice is systemic. That systemic inequity derives in no small part from there seldom being public evidence of the respective parts taken by collaborators in a particular research since the standard format of the scientific paper "presents an immaculate appearance that reproduces little or nothing of the intuitive leaps, false starts, mistakes, loose ends, and happy accidents that actually cluttered up the inquiry. *The public record of science therefore fails to provide many of the source materials needed to reconstruct the actual course of scientific developments*" (Merton 1968:4; italics inserted).[44]

Absent such detailed information, fellow scientists and scholars are evidently inclined to think it "reasonable" that the more accomplished collaborator with a history of major contributions to the field — i.e., the one with the far better "track record" as it is often put — has probably originated a joint work or contributed more to it, — unless there is compelling evidence to the contrary. This, even though such a probabilistic inference of course tells us next to nothing about the particular case with certainty.

However, in the case of the Thomas theorem, the compelling evidence is there and this time it is not to the contrary.

References

Bartlett, John. [1882] 1992. *Familiar Quotations,* edited by Justin Kaplan. 16th ed. Little, Brown.

Blumer, Herbert. 1939. *An Appraisal of Thomas and Znaniecki's 'The Polish Peasant in Europe and America.'* Social Science Research Council.

Broad, William, and Nicholas Wade. 1983. *Betrayers of the Truth.* Simon & Schuster.

Burchell, Howard B. 1985. "Thoughts on Eponyms." *International Journal of Cardiology* 8: 229-34.

Cole, Jonathan R. 1979. *Fair Science: Women in the Scientific Community.* Free Press.

Cole, Jonathan R., and Harriet Zuckerman. 1984. "The Productivity Puzzle: Persistence and Change in Patterns of Publication of Men and Women Scientists." Pp. 217-58 in *Advances in Motivation and Achievement.* edited by P. Maehr and M.W. Steinkamp. JAI Press.

Cole, Stephen. 1970. "Professional Standing and the Reception of Scientific Discoveries." *American Journal of Sociology* 76:286-306.

____. 1992. *Making Science: Between Nature and Society.* Harvard University Press.

Coleman, James S. 1990. *Foundations of Social Theory.* Harvard University Press.

Coleman, James S., Elihu Katz, and Herbert Menzel. 1966. *Medical Innovation.* Bobbs-Merrill.

Crane, Diana. 1972. *Invisible Colleges.* University of Chicago Press.

---

[44] It therefore seems only appropriate that a paper based on private as well as public knowledge of the respective roles of the distinguished pair of collaborators in this case should depart from the rigid pieties of the standard format.

Cronin, Blaise. 1984. *The Citation Process*. Taylor Graham.

Doyle, Arthur Conan [1894] 1953. "Silver Blaze" in *The Complete Sherlock Holmes*. Doubleday.

Epictetus, [c. 110] 1926-1928. *The Discourses and the Encheiridion*. Translated by W.A. Oldfeather. 2 vols. Putnam.

Epstein, Cynthia Fuchs. 1970. *Woman's Place: Options and Limits in Professional Careers*. University of California Press.

____. 1988. *Deceptive Distinctions: Sex, Gender, and the Social Order*. Yale University Press & Russell Sage Foundation.

Garfield, Eugene. [1975] 1977. "The 'Obliteration Phenomenon' in Science — and the Advantage of Bring Obliterated." Pp. 396-98 in Garfield, *Essays of an Information Scientist*. Vol. 2. ISI Press.

____. 1982. "More on the Ethics of Scientific Publication: Abuses of Authorship Attribution and Citation Amnesia Undermine the Reward System of Science." Pp. 621-26 in Garfield, *Essays of an Information Scientist*. Vol. 5. ISI Press.

____. 1994. "Commentary." *The Scientist*. 17 Oct., 1994.

____. 1995. "When to Cite." *Library Journal*. In press.

Garfield, Eugene, Irving R. Sher, and Richard J. Torpie. 1964. *The Use of Citation Data in Writing the History of Science*. ISI Monographs. Institute for Scientific Information.

Gaston, Jerry. 1973. *Originality and Competition in Science: A Study of the British High Energy Physics Community*. University of Chicago Press.

Hagstrom, Warren. 1974. "Competition in Science." *American Sociological Review* 39:1-18.

James, William. [1890] 1950. *The Principles of Psychology*. Dover.

Kaplan, Norman. 1965. *Science and Society*. Rand McNally.

Keller, Evelyn Fox. 1985. *Reflections on Gender and Science*. Yale University Press.

____. 1991. "The Wo/Man Scientist: Issues of Sex and Gender in the Pursuit of Science." Pp. 227-36 in *The Outer Circle: Women in the Scientific Community*, edited by Harriet Zuckerman, Jonathan R. Cole, and John T. Bruer. W.W. Norton.

Kurtz, Lester. 1984. *Evaluating Chicago Sociology: A Guide to the Literature with an Annotated Bibliography*. University of Chicago Press.

Kuznets, Simon, and Dorothy S. Thomas (eds.). 1959-64. *Population Redistribution and Economic Growth: United States, 1870-1950*. 3 vols. American Philosophical Society.

Lewontin, R.C., and J.L. Hubby. 1985. "Citation Classic." *Current Contents/Life Sciences* 43:16.

Mead, George H. 1936. *Movements of Thought in the Nineteenth Century*. University of Chicago Press.

Merton, Robert K. [1938] 1970. *Science, Technology, and Society in Seventeenth-Century England*. Howard Fertig.

____. [1938] 1973. "Science and the Social Order." Pp. 254-66 in Merton 1973. University of Chicago Press.

____. [1942] 1973. "The Normative Structure of Science." Pp. 267-78 in Merton 1973. University of Chicago Press.

____. [1948] 1982. "The Self-Fulfilling Prophecy." Pp. 248-67 in Merton 1982, edited by Aaron Rosenblatt and Thomas F. Gieryn. Abt Books.

____. 1948a. "Discrimination and the American Creed." Pp. 99-126 in *Discrimination and National Welfare*, edited by R.M. MacIver. Harper & Brothers.

____. [1961] 1973. "Singletons and Multiples in Scientific Discovery." Pp. 342-70 in Merton 1973. University of Chicago Press.

____. [1961a] 1973. "Social Conflict over Styles of Sociological Work." Pp. 47-69 in Merton 1973. University of Chicago Press.

____. [1963] 1973. "Multiple Discoveries as Strategic Research Site." Pp. 371-82 in Merton 1973. University of Chicago Press.

____. [1949, 1957] 1968. *Social Theory and Social Structure*. Free Press.

____. 1973. *The Sociology of Science.* University of Chicago Press.

____. 1976. "Social Knowledge and Public Policy." Pp. 156-79 in *Sociological Ambivalence.* Free Press.

____. 1980. "On the Oral Transmission of Knowledge." Pp. 1-25 in *Sociological Traditions from Generation to Generation,* edited by Robert K. Merton and Matilda White Riley. Ablex Publishing Corporation.

____. 1981. "Our Sociological Vernacular." Pp. 100-106 in Merton 1982. Abt Books.

____. 1982. *Social Research and the Practicing Professions.* Edited by Aaron Rosenblatt and Thomas F. Fieryn. Abt Books.

____. 1983. "Florian Znaniecki: A Short Reminiscence." *Journal of the History of the Behavioral Sciences* 10:123-26.

____. 1984. "Socially Expected Durations: A Case Study of Concept Formation in Sociology." Pp. 262-83 in *Conflict and Consensus: In Honor of Lewis A. Coser.* Edited by Walter W. Powell and Richard Robbins. Free Press.

____. 1987. "Three Fragments from a Sociologist's Notebooks: Establishing the Phenomenon, Specified Ignorance, and Strategic Research Materials." *Annual Review of Sociology* 13, 1-28.

____. 1988. "The Matthew Effect in Science, II: Cumulative Advantage and the Symbolism of Intellectual Property." *Isis* 79:606-23.

____. 1989. "Le molteplici origini e il carattere epiceno del termine inglese *Scientist.*" Pp. 279-93 in *Scientia: L'immagine e il mondo.* Scientia.

____. [1965] 1993. *On the Shoulders of Giants: A Shandean Postscript.* University of Chicago Press.

____. 1995. "Opportunity Structure: The Emergence, Diffusion, and Differentiation of a Sociological Concept, 1930s-1950s." Pp. 3-78 in *The Legacy of Anomie Theory.* Vol. 6 of *Advances in Criminological Theory,* edited by Freda Adler & William S. Laufer. Transaction Publishers.

____. 1996. "De-Gendering 'Man of Science': The Genesis and Epicene Character of the Word *Scientist.*" In *Sociological Visions,* edited by Kai Erikson. Yale University Press. In press.

Merton, Robert K. [and Harriet Zuckerman]. [1968] 1973. "The Matthew Effect in Science." Pp. 439-59 in Merton 1973. University of Chicago Press.

Mills, C. Wright. 1943. "The Professional Ideology of Social Pathologists." *American Journal of Sociology* 49:165-80.

Ogburn, William F. 1922. *Social Change.* B.W. Huebsch.

Ogburn, William F., and Dorothy Thomas. 1922. "Are Inventions Inevitable?: A Note on Social Evolution." *Political Science Quarterly* 37:83-98.

Palmore, Erdman. [1962] 1971. "Sociologists' Class Origins and Political Ideologies." Pp. 238-43 in *The Phenomenon of Sociology,* edited by Edward A. Tiryakian. Meredith Corporation.

Price, Derek J. de Solla. [1963] 1986. *Big Science, Little Science . . . and Beyond.* Columbia University Press.

Reinharz, Shulamit. 1989. "Teaching the History of Women in Sociology: Or Dorothy Swaine Thomas, Wasn't She the Woman Married to William I.?" *The American Sociologist,* Spring 1989:87-94.

Rogers, Everett M. [1962, 1971, 1983] 1995. *Diffusion of Innovations.* The Free Press.

Sarton, George. 1931. *Introduction to the History of Science.* Vol. 2 in Two Parts. Carnegie Institution of Washington.

Schopenhauer, Arthur. [1851] 1974 *Parerga and Paralipomena.* Translated by E.F.J. Payne. Clarendon Press, Oxford.

Schutz, Alfred. 1962. *Collected Papers.* Edited by Maurice Natanson. 2 vols. Martinus Nijhoff.

Science Watch. 1995. "Really Big Science: Multiauthor Papers Multiplying in the 1990s." *Science Watch* 6:1-2.

Sills, David L., and Robert K. Merton (eds.). 1991. *Social Science Quotations: Who Said What, When, and Where.* Macmillan.

Smith, R.S. 1993. "Dorothy Swaine Thomas and the 'Thomas Theorem'." Paper presented to the Eastern Sociological Society.

____. 1995. "Giving Credit Where Credit Is Due: Dorothy Swaine Thomas and the 'Thomas Theorem'." *The American Sociologist.* In press.

Stigler, George J. 1961. "The Economics of Information." *Journal of Political Economy* 69: 213-25.

Stigler, Stephen M. 1980. "Stigler's Law of Eponymy." Pp. 147-57 in *Science and Social Structure,* edited by Thomas F. Gieryn. *Transactions of the New York Academy of Sciences,* series II, Vol. 39.

Tate, Leland. B. 1974. "W.I. Thomas: Pioneer Sociologist and His Kin." In *Historical Sketches of Southwest Virginia.* Southwest Virginia Historical Society.

Thomas, Dorothy Swaine. 1938. *Research Memorandum on Migration Differentials.* Social Science Research Council.

____. 1952. "Experiences in Interdisciplinary Research." *American Sociological Review* 17:663-69.

Thomas, Dorothy Swaine and Associates. 1929. *Some New Techniques for Studying Social Behavior.* Teachers College, Columbia University.

Thomas, Dorothy Swaine, and Richard S. Nishimoto. 1946. *Japanese-American Evacuation and Resettlement: The Spoilage.* University of California Press.

Thomas, Dorothy Swaine, Charles Kikuchi, and James Sakoda. 1952. *Japanese-American Evacuation and Resettlement: The Salvage.* University of California Press.

Thomas, W.I. 1917. "The Persistence of Primary-group Norms in Present-day Society and Their Influence in Our Educational System." Pp. 159-97 in *Suggestions of Modern Science Concerning Education,* by Herbert S. Jennings, John B. Watson, Adolf Meyer, and W.I. Thomas. Macmillan

____. 1923. *The Unadjusted Girl.* Little, Brown.

____. 1927 "Situational Analysis: The Behavior Pattern and the Situation." *Publications of the American Sociological Society* 22:1-13.

____. 1931. "The Relation of Research to the Social Process." Pp. 175-94 in *Essays on Research in the Social Sciences.* The Brookings Institution.

____. 1937. *Primitive Behavior.* McGraw-Hill.

____. 1951. *Social Behavior and Personality: Contributions of W.I. Thomas to Theory and Social Research.* Edited by Edmund H. Volkart. Social Science Research Council.

____. 1966. *On Social Organization and Social Personality: Selected Papers.* Edited and with an introduction by Morris Janowitz. University of Chicago Press

Thomas, W.I., and Dorothy Swaine Thomas. 1928. *The Child in America: Behavior Problems and Programs.* Knopf.

Thomas, W.I., and Florian Znaniecki. [1918-20] 1927. *The Polish Peasant in Europe and America.* Five volumes in two. Knopf.

Vaihinger, Hans. [1911] 1924. *The Philosophy of 'As If.'* Routledge-Kegan Paul.

von Wiese, Leopold. 1932. *Systematic Sociology, On the Basis of the Beziehungslehre und Gebildelehre.* Adapted and amplified by Howard Becker. Wiley.

Waller, Willard W. 1932. *The Sociology of Teaching.* Wiley.

____. 1938. *The Family: A Dynamic Interpretation.* Cordon.

Young, Kimball. 1930. *Social Psychology: An Analysis of Social Behavior.* Crofts.

____. 1931. *Social Attitudes.* Henry Holt.

____. 1995. *On Sociology in Transition, 1912-1968: An Oral Account by the Thirty-fifth President of the ASA,* edited by Fred B. Lindstrom, Ronald A. Hardert, and Laura L. Johnson. University Press of America.

Zuckerman, Harriet. 1968. "Patterns of Name-Ordering Among Authors of Scientific Papers: A Study of Social Symbolism and its Ambiguity." *American Journal of Sociology* 74:276-91.

____. 1977. "Deviant Behavior and Social Control in Science." Pp. 87-137 in *Deviance and Social Change*, edited by E.A. Sagarin. Sage Publications.

____. 1978. "Theory Choice and Problem Choice in Science." *Sociological Inquiry* 48:65-95.

____. 1988a. "The Sociology of Science." Pp. 511-74 in *Handbook of Sociology*, edited by Neil Smelser. Sage Publications.

____. 1988b. "Intellectual Property and Diverse Rights of Ownership in Science." *Science, Technology, and Human Values.* Vol. 13, nos. 1 & 2:7-16.

____. 1989."Accumulation of Advantage and Disadvantage: The Theory and Its Intellectual Biography." Pp. 153-76 in *L'Opera di R.K. Merton e la sociologia contemporanea*, edited by Carlo Mongardini and Simonetta Tabboni. Edizioni Culturali Internazionali.

____. Jonathan R. Cole, and John T. Bruer (eds.). 1991. *The Outer Circle: Women in the Scientific Community*. W.W. Norton.

Zuckerman, Harriet, and Jonathan R. Cole. 1994. "Research Strategies in Science," *Creativity Research Journal* 7:391-405.

Zuckerman, Harriet, and Robert K. Merton. 1971. "Patterns of Evaluation in Science: Institutionalization, Structure and Functions of the Referee System." *Minerva* 9:66-100.

____. 1972. "Age, Aging, and Age Structure in Science." Pp. 292-356 in *A Sociology of Age Stratification*, edited by Matilda White Riley, Marilyn Johnson, and Anne Foner. Russell Sage Foundation.

The More, the Better? A Four-Nation Study of the Inclusion of Women in Symphony Orchestras*

JUTTA ALLMENDINGER, *Munich University*
J. RICHARD HACKMAN, *Harvard University*

Abstract

This article explores what happens when members of one identity group enter an elite institution that historically has been dominated by another. Specifically, we examine associations between the gender composition of professional symphony orchestras and several outcome measures — players' reports about orchestral functioning, the quality of members' relationships, and their own motivation and satisfaction. Analyses draw on quantitative, qualitative, and historical data collected for a comparative study of 78 symphony orchestras in four nations: the U.S., the United Kingdom, the former East Germany, and the former West Germany. Outcome measures decline as women's representation increases until the proportion of women approaches 50%. Then, the downward trend flattens or reverses for many variables. The dynamics of gender integration are found to differ (1) at low versus moderate levels of female representation, and (2) as a function of national context.

What happens as members of one identity group become an increasing presence in an organization or institution that has long been dominated by another — that is, when the history of identity group exclusivity is being altered? This question can be asked of blacks entering traditionally white organizations, and vice versa. It can be asked of Palestinians entering traditionally Israeli organizations, and vice versa. And it can be asked of women entering traditionally men's organizations, and vice versa.

* *Portions of the research reported here were financially supported by the Max Planck Institut für Bildungsforschung and by the Harvard Graduate School of Business Administration. We thank Larissa Kowal-Wolk, Erin Lehman, Rebecca Roters, and Jay Tucker for their invaluable assistance in the research on which this article is based. Helpful comments on a draft of the article were provided by Teresa Amabile, Mahzarin Banaji, Sandra Bem, Hannah Brückner, Alice Eagly, Adam Galinsky, Connie Gersick, Sue-Ellen Hershman, Rosabeth Kanter, Aage Sørensen, and Ruth Wageman. Direct correspondence to J. Richard Hackman, Department of Psychology, William James Hall, Harvard University, Cambridge, MA 02138.*

Our focus is on the movement of increasing numbers of women into professional symphony orchestras, relatively elite and traditionally male organizations whose work requires close coordination among members (Block & Neuls-Bates 1979; Lehmann 1982). As will be seen, the gender integration of professional symphony orchestras is still in its early stages — which provides a historically unique window of opportunity to examine gender recomposition processes. We address what is happening at three levels of analysis: (1) to the *individuals* involved — the old hands as well as the new entrants, (2) to the *relationships* among members, and (3) to the structures, processes, and performance of the *organization as a whole.*

Existing literature on group and organizational composition provides some clues about what to expect as the proportion of women in orchestras increases. We know, for example, that the characteristic interaction styles of women and men differ (Dion 1985; Eagly 1987; Moreland & Levine 1992; Wood & Rhodes 1992) and that gender-homogeneous groups often have "smoother" processes — i.e., less conflict, more predictability, and less member anxiety — than heterogeneous groups (Jackson 1992; Morgan & Lassiter 1992; Pfeffer 1983). Smooth processes are not necessarily advantageous, however. The gender-characteristic norms that tend to emerge in homogeneous groups can limit both the personal learning of members and their collective task performance (Martin 1985:332). Moreover, conflict prompted by diversity and moderate deviation from traditional norms can enhance group creativity — a matter of special importance to performing groups such as symphony orchestras (Hackman 1992; Hoffman 1965; Nemeth & Staw 1989).

Longitudinal research exploring the implications of these phenomena for organizations that are in the midst of demographic transition would be ideal, but such studies have not yet been done. Still, by examining life and work in organizations that vary widely in how far they have come toward relative gender balance, insight can be obtained about the dynamics and effects of such transitions. The present research addresses three questions in this regard. The first assesses the size and direction of gender composition effects; the second and third begin to explore the mechanisms by which such effects come to pass.

(1) What happens to life and work in symphony orchestras as the proportion of women increases from token levels to a near-majority? A great deal is known about gender-homogeneous social systems (especially those that are dominantly male) and about gender dynamics when women have but token representation. (For an interpretive review of research on tokenism, see Yoder 1991.) But how do individual perceptions and attitudes, player relationships, and organizational functioning change as the proportion of minority members becomes substantial?

As noted by Martin (1985) and Yoder (1991), a number of scholars have suggested that once the proportion of women increases beyond the token level, life at work should improve — at least for women, and perhaps for their male colleagues as well. Evidence for this proposition is shaky at best (Blum & Smith 1988). Few studies of gender composition in organizations have sampled female representation widely enough, or included a broad enough set of dependent variables, to reliably track what happens as an organization evolves from a state of male exclusivity to relative gender balance.

Findings from those few studies whose designs do permit inferences to be made about the effects of increased female representation provide little basis for optimism. South et al. (1982), for example, examined the full spectrum of female representation across 76 work groups of varying size in six offices of a state agency (mean percentage of women = 51%, s = .27) and found that the higher the proportion of women, the less social support and encouragement for promotion they received. Two studies assessed the effects of gender composition in universities — which, like symphony orchestras, are relatively elite institutions. Toren (1990) found that the greater the proportion of women on the faculties of Israeli universities, the lower their academic ranks relative to those of their male colleagues. Alexander and Thoits (1985) assessed the academic performance of female students who recently had entered a traditionally male institution; they found evidence for underachievement by those token females who were enrolled in dominantly male departments.

A number of studies have shown that disparities in income between numerical minorities and majorities become greater as the relative size of the minority group increases (e.g., Frisbie & Neidert 1976; Martin & Harkreader 1993). Especially informative is a longitudinal study by Pfeffer and Davis-Blake (1987) that examined the impact of changes in gender composition on the salaries of administrators at more than 800 colleges and universities. They found a monotonic decline in the salaries of both male and female administrators as the proportion of women increased from token levels to the 30-40% range. Beyond that "tipping point," however, further increases in the proportion of women made little or no difference in salaries.

In the present research, we assess whether the negative consequences for individuals that sometimes accompany the entry of women into traditionally male organizations also hold for variables at the group and organizational levels of analysis. By collecting data from male and female members of orchestras that are spread across a wide portion of the gender composition spectrum in four nations, we should be able to reliably assess both the magnitude and the direction of gender composition effects. Moreover, we can determine if there is, for variables other than salary, evidence of the tipping point identified by Pfeffer and Davis-Blake.

Research Question 1a: How strong are gender composition effects on the perceptions and experiences of orchestra members? Are composition effects stronger or weaker than those obtained for comparison variables at other levels of analysis — specifically, player gender at the individual level and nation at the contextual level?

Research Question 1b: What is the direction of gender composition effects? That is, does the proportion of women in an orchestra vary directly or inversely with measures of individual well-being, relationship quality, and organizational vitality?

Research Question 1c: At what point on the gender composition continuum, if any, is a tipping point observed — and for what measures is it found?

(2) *Do female and male members respond differently to greater female representation in orchestras?* It may be that any main effects of gender composition are due primarily to the reactions of the new entrants — or, perhaps, to those of the old hands. Or it may be that members of both groups react similarly to changes in the proportion of female players.

Females. We know that highly skewed composition, which is inevitable when nontraditional members begin to enter a formerly exclusive group or organization, presents significant personal challenges to members of the entering group. The classic empirical study documenting the difficulties that women face when they enter predominantly male organizations is Kanter (1977a). This work explored in detail the experiences of women who constituted a small but growing minority in a large male-dominated firm.

According to Kanter (1977b), three processes operate in parallel to generate the unsatisfactory experiences of minority individuals. The first is heightened *visibility*, which results in minority members receiving a disproportionate share of the attention of other system members; that, in turn, leads to their experiencing excessive pressure to perform well. The second is increased *polarization* between the majority and minority groups, which results in exaggeration of those attributes that differentiate the minority members from the majority. The third is *assimilation*, which involves distortion of minority members' characteristics to fit with majority members' preexisting generalizations about them; this makes it difficult for minority members to escape their traditional roles.[1]

In the present context, Kanter's findings suggest that any dysfunctions associated with the entry of women into symphony orchestras should initially be more pronounced for the entering females than for the veteran male players. As the number of female players increases, however, women could increasingly provide both support and reality checks for one another — thereby ameliorating some of the difficulties encountered by the "pioneer" females. Indeed, in orchestras that are relatively gender balanced, women's reports about orchestral life might well be indistinguishable from those of their male colleagues.

Males. It may be that main effects of gender composition are due as much, or more, to the perceptions and experiences of the veteran men than to those of the entering women (Blalock 1967). Life in a homogeneously male orchestra surely is not much affected by the presence of one or two women, such as a female harp player. As their numbers increase, however, women become, in Yoder's (1991) phrase, an "intrusive" presence on high-status turf that previously has been an exclusively male province.

The entry of substantial numbers of women into orchestras may raise concerns among men about their dominance of the organization and control of its operations. Moreover, the presence of a noticeable and growing group of women may increase the salience of men's own identity group and prompt concerns about women taking over jobs or resources that historically have been accessible only to them (Cockburn 1985, chap. 8; Tolbert, Simons & Oberfield 1994). Men also may worry that the high status of their profession will be compromised by the entry of women — in effect, that their professional currency will be devalued (Martin 1985; Reskin & Roos 1990; Touhey 1974). Finally, the gender-characteristic patterns of interaction that the male members are likely to have evolved may now be challenged by new ways of thinking and acting exhibited by people who are perceived to be "not like us."[2]

Both Genders. The minority processes posited by Kanter and the majority processes posited by Blalock generate contrasting predictions about what will happen as the proportion of females in traditionally male organizations increases (Jacobs 1992; Pfeffer & Davis-Blake 1987; Tolbert, Simons & Oberfield

1994). The Kanter position is that things will get better, mainly because of what the women are doing; the Blalock position is that things will get worse, mainly because of how the men are reacting.

There remains a third possibility — namely, that gender composition effects will be found for *both* female and male members. This could occur (1) because the separate processes described above for females and males are both operating (and, perhaps, reinforcing one another), or (2) because orchestras with mixed gender composition actually do operate less well — and that fact is noted and reported upon by all members regardless of gender.

The relative validity of these three alternatives — that is, that gender composition effects are due primarily to the responses of female players, to those of male players, or to those of all orchestra members — can be assessed by examining the strength and direction of any interactions that are found between individual gender and orchestra gender composition in shaping players' reports about their orchestras.

Research Question 2a: Do women and men respond differently to different gender composition levels? Or do all members react similarly as orchestras become less exclusively male?

Research Question 2b: Which gender exhibits the more pronounced change in perceptions and experiences as gender composition becomes more balanced?

(3) To what extent do gender composition effects differ as a function of national context? Specifically, do the effects of women entering traditionally male organizations differ for nations that encourage women's full participation in the work force, compared to nations that are relatively neutral to (or actively discouraging of) women's participation?

A country's policies, practices, and social structures can powerfully influence the receptivity of the workplace to women (Hofstede 1980). Indeed, Schmidt (1992:47) concluded, based on data from a comparative study of gendered labor force participation in the OECD nations, that "the large cross-country variation in labor force participation rates of men and women can, indeed, be properly appreciated only in the context of 'deep-seated institutional differences' between the countries studied."

Organization members operate simultaneously in multiple gender contexts (Martin & Harkreader 1993). Between-nation differences in receptivity to women's participation in the national work force, therefore, may affect not just the overall level of women's work force participation but also how changes in organization-level gender composition unfold.

Research Question 3: How do gender composition dynamics differ for countries that have relatively high versus relatively low rates of women's participation in the national work force?

In sum, the present research explores the main and interactive effects of the proportion of women in orchestras, individual gender, and national context in predicting (1) individual motivation and satisfaction, (2) the quality of relationships among players, and (3) their reports about how, and how well, their orchestras function. Simultaneously considering variables at the individual, organization, and contextual levels of analysis, we believe, can yield insight into

the dynamics of gender compositional effects that would escape detection in research conducted at any single level of analysis.

Research Setting

The present research draws upon data from 78 professional symphony orchestras in four nations: the U.S., the United Kingdom, the former East Germany, and the former West Germany. Symphony orchestras are defined as ensembles whose primary purpose is public performance of those orchestral works generally considered to fall within the standard symphonic repertoire, and whose members are compensated nontrivially for their services. Such orchestras have several unique features that make them advantageous for examining the research questions specified above.

SALIENCE OF COMPOSITION DYNAMICS

Gender composition dynamics are likely to be especially evident and consequential for players in symphony orchestras, for two reasons. First, the work of orchestras is done literally in concert and requires a very high level of between-member coordination. In contrast to some other types of organizations (e.g., field sales operations, or schools in which teachers have relatively autonomous responsibility for their classrooms), any change in orchestra composition will be experienced immediately and continuously by all members. Second, orchestras are highly visible and relatively elite organizations. There are many qualified applicants for each opening in a professional symphony orchestra, and musicians who do secure orchestral positions enjoy relatively high status in their profession and their communities. Therefore, any change in orchestra demography is likely to be both salient and significant for all members.

ORCHESTRA TASK AND MEMBER TALENT

Variation associated with both organizational tasks and member talent is relatively low across both orchestras and nations. The core task of symphony orchestras is both well-defined and similar across nations. Although there are indeed large cultural and contextual differences among nations, symphony orchestras around the world play largely the same repertoire with roughly the same number and mix of players.[3] Thus, by defining the domain of the research to include *only* orchestras that perform the standard symphonic repertoire, most task-based variance is automatically controlled.

Use of professional symphony orchestras as the research setting also automatically controls a great deal of variation in member talent. Because there are far more talented musicians looking for work than there are open positions, most orchestras are staffed with fully qualified players. By studying symphony orchestras, then, it is possible to avoid the problem, common to comparative organizational research, that differences in organizational tasks or in the mean skill levels of organization members are confounded with organization-level variables or with the properties of national contexts.

RATES OF WOMEN'S PARTICIPATION

There is substantial variation both within and between nations in orchestra gender composition. This makes it possible to examine compositional main effects across a relatively large portion of the gender composition spectrum as well as to detect any interactions between orchestra composition and national context that shape the perceptions and experiences of orchestra members.

Although women historically have been active in music as teachers, singers, and amateur performers (Edwards 1991), it is only in recent decades that they have become accepted as full members of professional symphony orchestras. In the U.S., for example, it was not until the mid-1950s that women began to be accepted into major orchestras in more than token numbers; in the two German nations, that did not occur until the 1980s (Allmendinger & Hackman 1993a). As recently as the early 1990s, there still were relatively few female players in professional symphony orchestras, as is evident in the left panel of Figure 1. Of the 78 orchestras in our sample, only eight employ women and men in roughly equal numbers; in more than half of the orchestras, fewer than 25% of the players are female. Nonetheless, the range of female representation in the sample as a whole (2% to 59%, median = 21%) is satisfactory for our research purposes.

The four countries in our sample differ substantially in women's work force participation. Data from the International Labor Organization show that the proportion of working women in 1989 (when we drew our sample of orchestras) ranged from around 40% in West Germany and the United Kingdom to 45% in the U.S. and 50% in East Germany. Since 1960, women's workforce participation has increased in all four nations, most sharply in the U.S. (a gain of 13%), moderately in the U.K. and East Germany (gains of 8% and 6% respectively), and least in West Germany (a 3% gain).

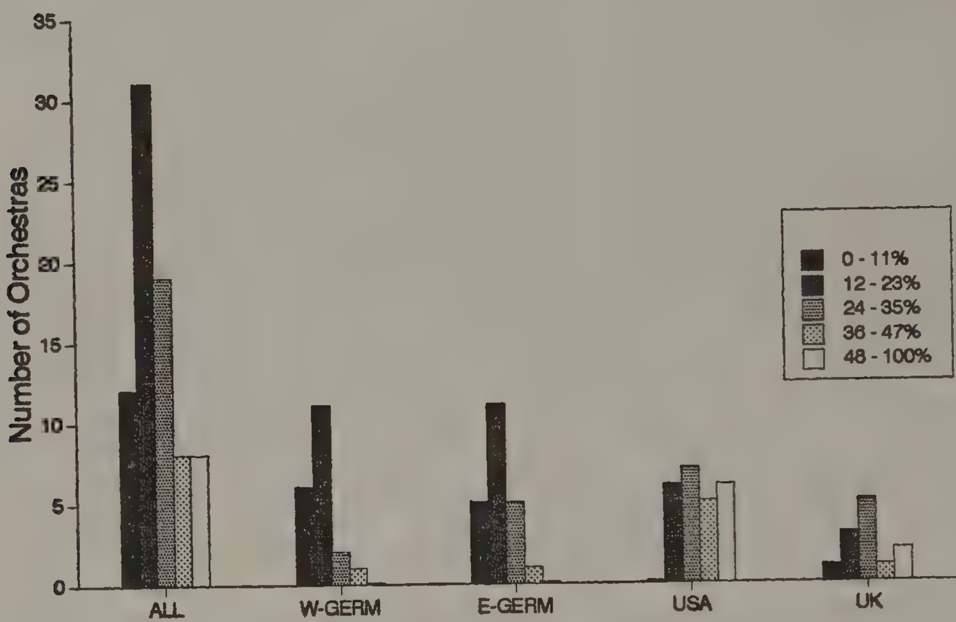
The distribution of women's representation in orchestras in each of the four countries is shown in Figure 1. The U.S. has the highest proportion of women (median = 36%), which is the lower bound of Kanter's (1977b) "low balanced" category.[4] Next comes the United Kingdom, which, with a median of 30% women, falls into Kanter's "high tilted" category. Finally, the median proportion of women in both West and East Germany is 16%, placing both of these nations in her "low tilted" category.

There are many differences among the nations in our sample, of which rates of women's participation in the work force is but one. Any attributions about the reasons for empirical findings involving differences among nations must therefore be made cautiously and conservatively. We will see, nonetheless, that differences in national context do play a significant role in how recomposition processes are unfolding in symphony orchestras in the four nations studied.

Method

This research draws upon archival, observational, interview, and survey data collected for a four-nation comparative study of leadership and mobility in symphony orchestras. We present here only those aspects of the methodology that are germane to present purposes; for a detailed description of project methods, see Allmendinger et al. (1993).

FIGURE 1: Women's Representation in Orchestras in 1990



SELECTION OF ORCHESTRAS

Professional symphony orchestras in all four countries announce all vacancies for players in union or government publications. It therefore was possible to enumerate the population of symphony orchestras in each country by listing all orchestras that announced at least one open player position between 1985 and 1989. We obtained 1989 budgets for each of these orchestras from industry, government, and trade union reports and used the distribution of salary budgets to select approximately equal numbers of relatively affluent and widely known "major" orchestras and generally smaller, more regionally oriented "minor" orchestras.

Major orchestras were randomly drawn from the top decile (or, in the United Kingdom, which had fewer orchestras, quintile) of the distribution; minor orchestras were randomly drawn from each of the remaining deciles (or quintiles) until a sample of about 20 orchestras from each country was obtained.[5] Of the 81 orchestras invited to participate, 78 agreed (96%) (the exceptions were one minor orchestra in the U.S. and one major and one minor orchestra in the United Kingdom).

SOURCES OF DATA

Data were collected during 1990 and 1991. We summarize below procedures and instruments that were common across orchestras in all four countries; special procedures that were required for data collection in East Germany are described by Allmendinger and Hackman (1992).

Four types of data were obtained for each orchestra: archival information, interviews, observations, and player surveys. Archival information includes data in the public domain (such as published histories and newspaper or magazine articles) and information obtained from an orchestra's files (such as concert programs, rosters of players and managers, financial data, and attendance figures). After reviewing published information about an orchestra, one or more project researchers visited it to conduct interviews, observe the orchestra, and collect data from the orchestra's files.

We also sought permission from orchestra management (and, when appropriate, from the players' orchestra committee) to administer a questionnaire to a sample of players. Surveys were anonymous, although they did identify the respondent's orchestra, section, and status as a principal or as a tutti player. They usually were distributed by mail (and always with a preaddressed and stamped return envelope) to a stratified random sample of 15 players in each orchestra: two first violins, two second violins, one viola, one cello, one bass, two woodwinds, two brass, one percussion, and three principal players (one each from the strings, woodwinds, and brass/percussion). After the first round of surveys had been returned, additional surveys were sent to most orchestras. The same sampling plan was used in the second round, except that women were deliberately oversampled to obtain enough responses from female players to permit comparative analyses between the responses of men and those of women.

Six participating orchestras chose not to have surveys distributed to their players: four in West Germany and two in the U.S. The sample for the main analyses reported in this article is 72 orchestras and 992 respondents. The number of orchestras, number of respondents, and survey response rates for each country are as follows. East Germany: 22 orchestras, 357 respondents (59%); West Germany: 16 orchestras, 186 respondents (45%); U.S.: 22 orchestras, 290 respondents (52%); United Kingdom: 12 orchestras, 159 respondents (51%).

SURVEY MEASURES

The player survey consisted of 87 items distributed across five domains. Originally drafted in English, the survey was translated into German (and partially back-translated) by bilingual members of the research team. Sixteen composite measures were derived, based on the *a priori* framework used in constructing the survey supplemented by principal component analyses and inspection of item means, standard deviations, and interitem correlations. Six single items of special interest also were retained, resulting in a total of 22 measures. All measures range from 1 (low) to 7 (high) except for those assessing the behavior of the music director, which range from 1 to 5.

Means, standard deviations, and internal consistency reliabilities (assessed using Chronbach's alpha) for all measures are provided in Table 1; descriptions of the measures are provided in the Appendix. Means and standard deviations are satisfactory, although internal work motivation approaches the scale ceiling. Internal consistency reliabilities range from .58 to .92 (median = .76). Two measures have marginal reliabilities: satisfaction with work relationships (.58) and adequacy of resources (.59).

Discriminant validities of the measures were assessed by comparing the average intercorrelation among items that comprise a given composite score with those items in the same section of the survey that do *not* contribute to that score ("off-diagonal" correlations). The median within-composite correlation is .46, compared to a median off-diagonal correlation of .24, which is satisfactory. Three measures, however, do have relatively weak discriminant validities: integrity as an ensemble, music director organizational initiatives, and music director coaching.

Finally, within-composite correlations, off-diagonal correlations, and intercorrelations among the 16 composite measures were compared across countries. Patterns of intercorrelations were found to be quite similar across nations, which provides reassurance about the comparability of players' responses to the English and German versions of the survey and permits use of the same composite scores for all countries. (For details, see Allmendinger et al. 1993).

ANALYTIC STRATEGY

The primary data analysis was accomplished using the SAS General Linear Models program. Each dependent variable was predicted from gender, proportion of women, country, and all interactions among these factors. A hierarchical model was used, in which gender (the design factor at the individual level of analysis) was entered first, followed in turn by proportion of women (the organizational level), country (the contextual level), and all two- and three-way interactions among these factors. To account for orchestra-specific variance, the model also included, as the last term, orchestra (a random effect) nested within standing (i.e., major versus minor orchestras).

Because the proportion of women and men in a given orchestra who returned surveys did not match the actual proportion of female and male members of that orchestra, weights were applied to the data from female respondents in each orchestra so that their contributions to orchestra-level means corresponded to their actual representation in that orchestra. Also, the uneven distribution of orchestras across design factors made it impossible to use least-square procedures to estimate some cell means. Therefore, means reported in this article were computed by first averaging the responses of all orchestra members and then averaging these orchestra-level means across the design factors of interest.

TABLE 1: Means and Standard Deviations of Survey Measures[a]

	Items	Mean	Std. Dev.	Reliability
Organizational features of the orchestra				
Integrity as an ensemble	8	5.05	1.00	.73
Orchestra structure: task and composition	5	4.86	1.03	.65
Player involvement and opportunities	8	4.10	1.11	.76
Adequacy of resources	2	4.20	1.81	.59
Single items:				
Recognition (excellent playing pays off)		5.50	1.51	—
Recruitment (fair and effective)		5.28	1.60	—
Behavior of the music director				
Direction-setting	5	3.67	1.09	.92
Organizational initiatives	7	3.15	0.98	.84
Coaching	5	2.82	0.90	.79
Single item:				
Music director authority		3.74	1.92	—
Orchestral processes				
Quality of relationships	7	4.76	1.03	.78
Musical outcomes	3	4.81	1.25	.70
Single items:				
Finances (getting stronger financially)		3.35	1.84	—
Stability (organization not falling apart)		5.92	1.54	—
Player motivation and satisfaction				
General satisfaction	3	5.61	1.21	.74
Internal work motivation	3	6.21	0.79	.69
Job involvement	1	4.70	2.04	—
Specific satisfactions				
Compensation	2	3.87	1.75	.73
Job security	2	4.82	1.48	.88
Management	3	4.28	1.48	.86
Work relationships	3	4.93	0.98	.58
Growth opportunities	4	4.84	1.17	.79

[a] Number of responses ranges from 886 to 924.

Results

Findings are presented separately for each of the three research questions set forth in the introduction to this article. First we examine the size and direction of gender composition main effects; then we examine how gender composition dynamics differ (1) for female and male orchestra members, and (2) as a function of national context.

GENDER COMPOSITION

Table 2 shows the size and statistical reliability of all main effects and two-way interactions for each of the five groups of dependent variables: perceived organizational features, music director behaviors, orchestral processes, overall player motivation and satisfaction, and player satisfaction with specific aspects of their work lives.

Effect Sizes

Gender composition is significantly related to 19 of the 22 dependent measures, most strongly with measures of perceived organizational features and least strongly with assessments of music director behavior. Of the two comparison variables, country and respondent gender, the former also is strongly related to the dependent measures (21 of 22 associations significant), but the latter is not (only 5 associations significant).

The numerous and often substantial main effects for country are the subject of a forthcoming report and therefore not discussed here. Our present interest, addressed in a subsequent section, is in *interactions* involving nation — i.e., how national context alters gender composition dynamics.

The findings regarding individual gender, on the other hand, merit close inspection. Whether as a main effect or in interaction with gender composition or with country, individual gender controls little variation in our measures. Across all dependent variables, the median proportion of variance controlled by gender and its interactions is less than 1% (compared to a median of 5% for both proportion of women and country), even though gender was entered as the first term in the hierarchical data analysis model.

It is, perhaps, not surprising that there are few significant gender effects for the measures of organizational features. It is, after all, the same organization that men and women are describing, so the absence of strong gender effects is reasonable. The only statistically reliable gender effects for these measures are that men view their orchestras as better structured and as having greater integrity as ensembles than do women (F ratios of 7.1 and 22.3 respectively, $p < .01$ for both variables).

It is noteworthy, on the other hand, that there are few significant differences in how women and men describe the behavior of orchestra leaders, internal orchestra processes, and their own motivation and satisfaction. The only significant effects are that (1) women view the music director as more active in coaching players and in doing organizational work than do men ($F = 5.8$, $p < .05$ for coaching; $F = 6.0$, $p < .01$ for organizational initiatives), (2) women are more optimistic than men about how their orchestras are doing financially ($F = 23.4$,

TABLE 2: Size of Effects of Gender, Proportion of Women, and Country[a]

	Gender		Proportion of Women		Country		Gender by Proportion		Gender by Country		Proportion by Country	
	Sig.	%†	Sig.	%	Sig.	%	Sig.	%	Sig.	%	Sig.	%
Organizational features (6 measures)	2/6	0	6/6	8	6/6	4	5/6	1	1/6	0	6/6	4
Music director behavior (4 measures)	1/4	0	1/4	1	4/4	6	1/4	0	0/4	0	4/4	4
Orchestral processes (4 measures)	1/4	0	4/4	6	4/4	5	3/4	2	0/4	0	3/4	2
Overall player motivation and satisfaction (3 measures)	1/3	1	3/3	5	3/3	15	0/3	1	0/3	0	0/3	1
Satisfaction with specific organizational features (5 measures)	0/5	0	5/5	2	4/5	6	1/5	0	0/5	0	3/5	2

[a] Significant three-way interactions were obtained for 2 of the 22 variables. These interactions control less than 2% of the dependent variable variance, and neither is of substantive interest. They are not discussed further.

* The number of measures for which the effect is significant at $p = .01$.

† The mean percentage of the total variation of scale scores that is controlled by the effect.

$p < .01$), and (3) men are more involved with their work (and are happier about their opportunities for personal and professional growth) than are women ($F = 40.3$, $p < .01$ for job involvement; $F = 5.4$, $p < .05$ for growth satisfaction). Other than these differences, the perceptions and experiences of male and female players within a given orchestra are, on the whole, quite similar.

Direction of Effects

Means for all dependent measures, arranged by proportion of women, are presented in Table 3. Almost all the dependent measures decline significantly as the proportion of women in an orchestra increases from token to more substantial levels. However, 10 of the 19 significant effects exhibit a statistically significant reversal of the decline at some point on the gender composition spectrum (see the *F* ratios for the quadratic component in the far right column of the table).

TABLE 3: Dependent Measure Means by Orchestra Gender Composition

	Proportion of Women (%)					F ratio† (Main Effects)	Percent Var.	F ratio (Quad. Comp.)
	0-11	12-23	24-35	36-48	49-60			
Organizational features								
Integrity as an ensemble	5.63	5.32	4.92	4.64	4.42	43.5**	13	0.23
Orchestra structure	5.46	4.94	4.62	4.51	4.85	21.9**	7	10.35**
Player involvement	4.58	4.23	3.98	3.79	3.87	13.0**	5	0.32
Adequacy of resources	4.84	4.04	3.50	4.12	4.46	20.2**	6	23.27**
Recognition for good work	6.32	5.73	5.29	5.04	4.72	27.9**	10	1.15
Fair/effective recruitment	5.96	5.52	5.20	4.94	4.84	12.8**	5	0.24
Behavior of music director								
Direction-setting	3.84	3.73	3.60	3.76	3.46	3.4*	1	0.32
Organizational initiatives	3.19	3.15	3.11	3.30	3.10	1.5	1	...
Coaching	2.89	2.80	2.76	2.86	2.96	1.5	0	...
Music director authority	3.43	3.69	3.75	4.15	3.77	2.0	1	...
Orchestral processes								
Relationship quality	5.31	4.76	4.48	4.68	4.84	16.8**	6	6.28**
Musical outcomes	4.79	4.56	4.94	5.29	5.26	13.9**	5	4.88*
Improved finances	3.04	2.86	3.44	3.66	4.67	29.5**	9	7.12**
Organizational stability	6.56	6.02	5.64	5.31	6.02	15.6**	6	5.73*
Player motivation and satisfaction								
General satisfaction	6.20	5.73	5.35	5.45	5.39	15.1**	5	4.88*
Internal work motivation	6.29	6.26	6.09	6.35	6.11	4.2*	2	1.34
Job involvement	5.85	5.33	4.24	4.20	3.44	56.2**	8	2.74
Satisfaction with								
compensation	4.20	3.90	3.76	3.25	3.95	7.7**	2	1.82
job security	5.41	4.81	4.50	4.32	4.96	16.2**	5	11.28**
management	4.73	4.35	4.06	4.16	4.36	5.7**	2	3.39
work relationships	5.13	4.89	4.91	4.89	5.16	3.3*	1	4.89*
growth opportunities	5.22	4.87	4.63	4.80	4.98	6.1**	2	5.53*

† F ratios, p values, and percent variance figures are based on analysis of 992 respondents using the general linear model described in the text.

* $p < .01$ ** $p < .001$

All six measures of members' perceptions of their orchestras' organizational features show a significant decline as the proportion of women increases. The decline is monotonic for four of the six measures; for the other two (organizational structure and adequacy of resources) there is an initial decline followed by a significant upturn as the proportion of women becomes larger.

Reports about the behavior of the music director are generally unaffected by organizational gender composition. The measures of orchestral processes, however, are strongly and significantly related to the proportion of women in the orchestra. All four of these measures exhibit an initial decline followed by an upturn.

Finally, all measures of player satisfaction and motivation are significantly associated with gender composition, and all but two exhibit the now-familiar pattern: an initial decline followed by an upturn. The two exceptions are job involvement (which exhibits a monotonic decline) and internal work motivation (which shows no interpretable pattern).

To show the consistency of the emergent pattern across domains of dependent measures, we present in Figure 2 the means for six representative measures — two organizational features (integrity as an ensemble and organizational structure), two organizational processes (quality of interpersonal relationships and organizational stability), and two attitudinal measures (job involvement and satisfaction with job security).

Where the Variance Lives

Far more variation in our measures was found to be controlled by contextual and organizational factors than by gender assessed as an individual-level attribute. That finding is consistent with the increasing recognition of the importance of social and structural factors in explaining even those male-female differences that may initially appear to be native (Eagly 1987; Faludi 1991). Still, the nonexperimental character of our research design requires that special care be taken to ensure that the direction of our findings cannot be explained by other variables that covary with orchestra gender composition. In particular, three rival explanations must be considered: respondent age and experience, respondent instrument, and orchestra standing.[6]

Age and Experience. On average, women players in professional symphonies have served in their orchestras for substantially fewer years than have their male colleagues (means of 11 and 16 years, respectively) and they are, on the whole, considerably younger than their male colleagues. It is commonly found that older employees with more years of organizational experience are more satisfied with their work lives than are their younger colleagues (e.g., Hackman & Oldham 1976). We find the same phenomenon. Players in our sample over age 40 were significantly more satisfied than their younger colleagues on six of the eight measures of motivation and satisfaction. (The exceptions were satisfaction with job security and with pay.) This finding raises the possibility that the effects we are attributing to gender composition actually reflect the fact that orchestras with larger numbers of women are also, on the whole, younger orchestras. We tested this possibility by examining the relationship between gender composition and the six representative measures in Figure 2 separately for younger and older players. The findings, shown in Figure 3, replicate for

FIGURE 2: Effects of Proportion of Women

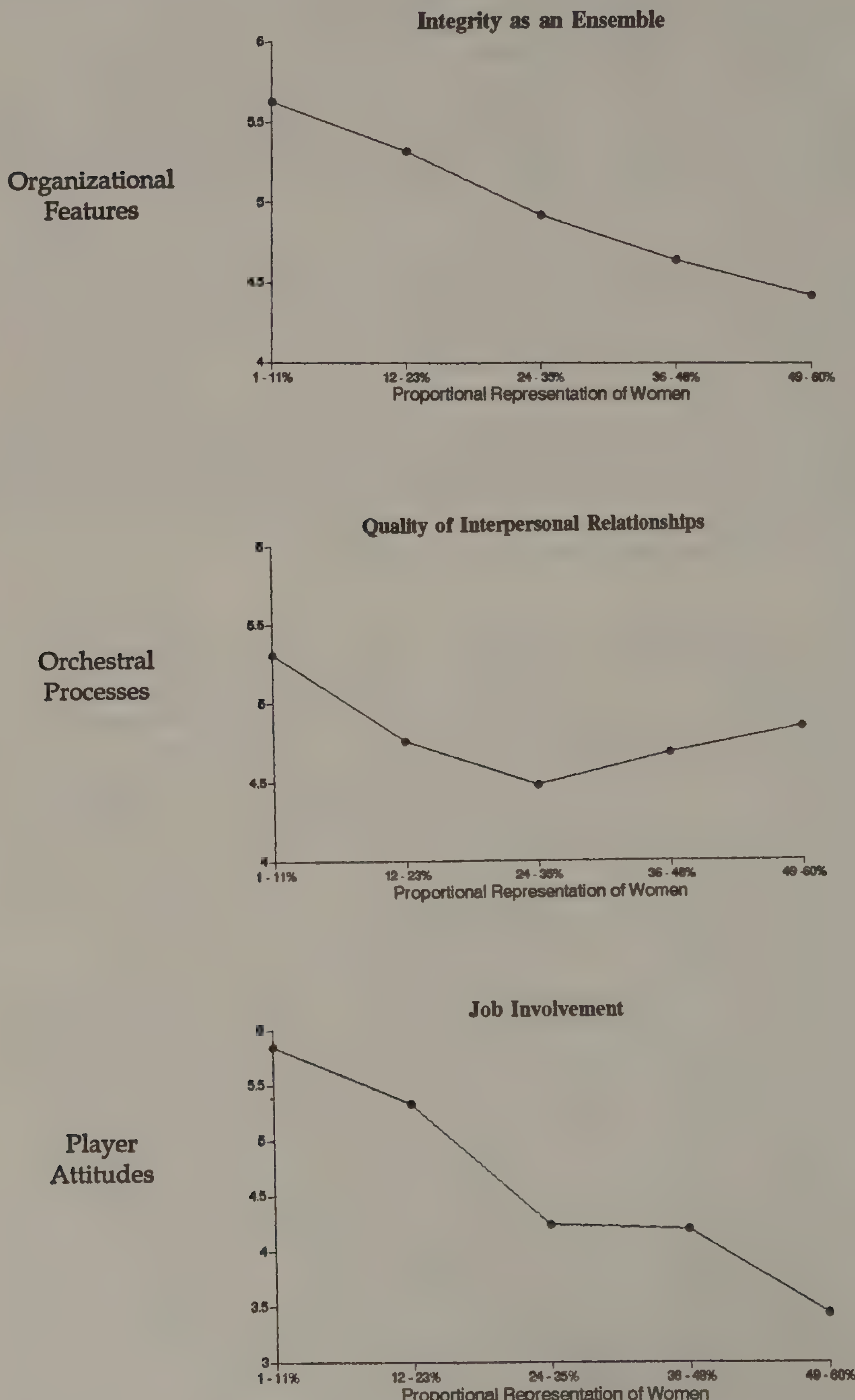
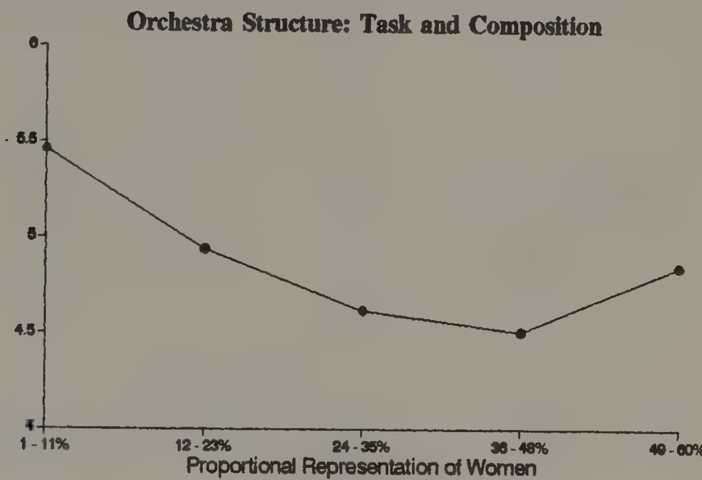
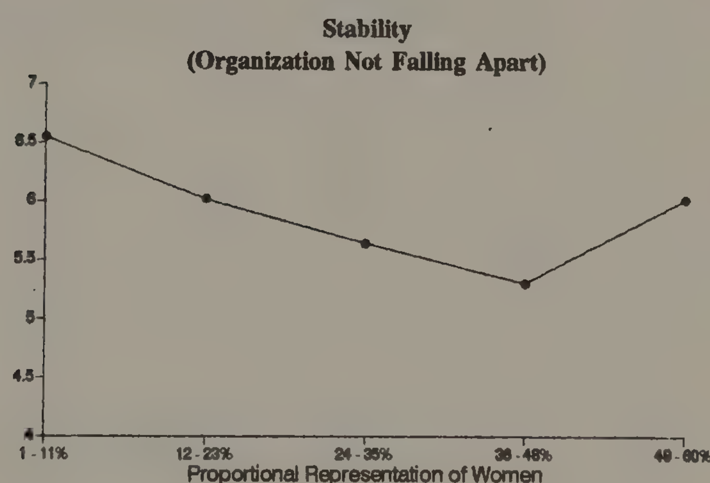


FIGURE 2: Effects of Proportion of Women (Continued)

Organizational Features



Orchestral Processes



Player Attitudes

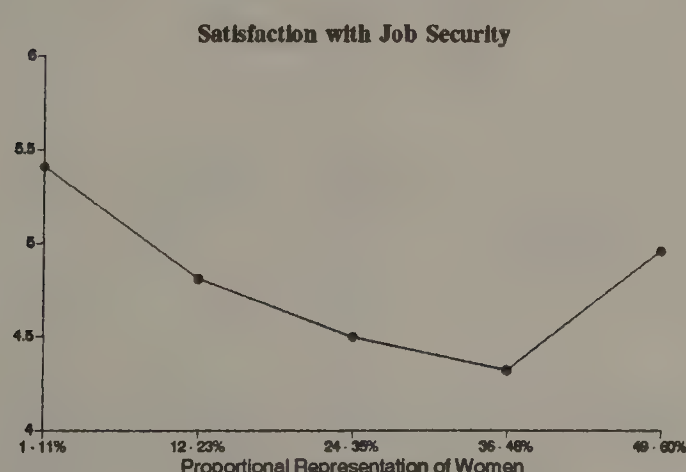


FIGURE 3: Proportion Effects for Younger vs. Older Players

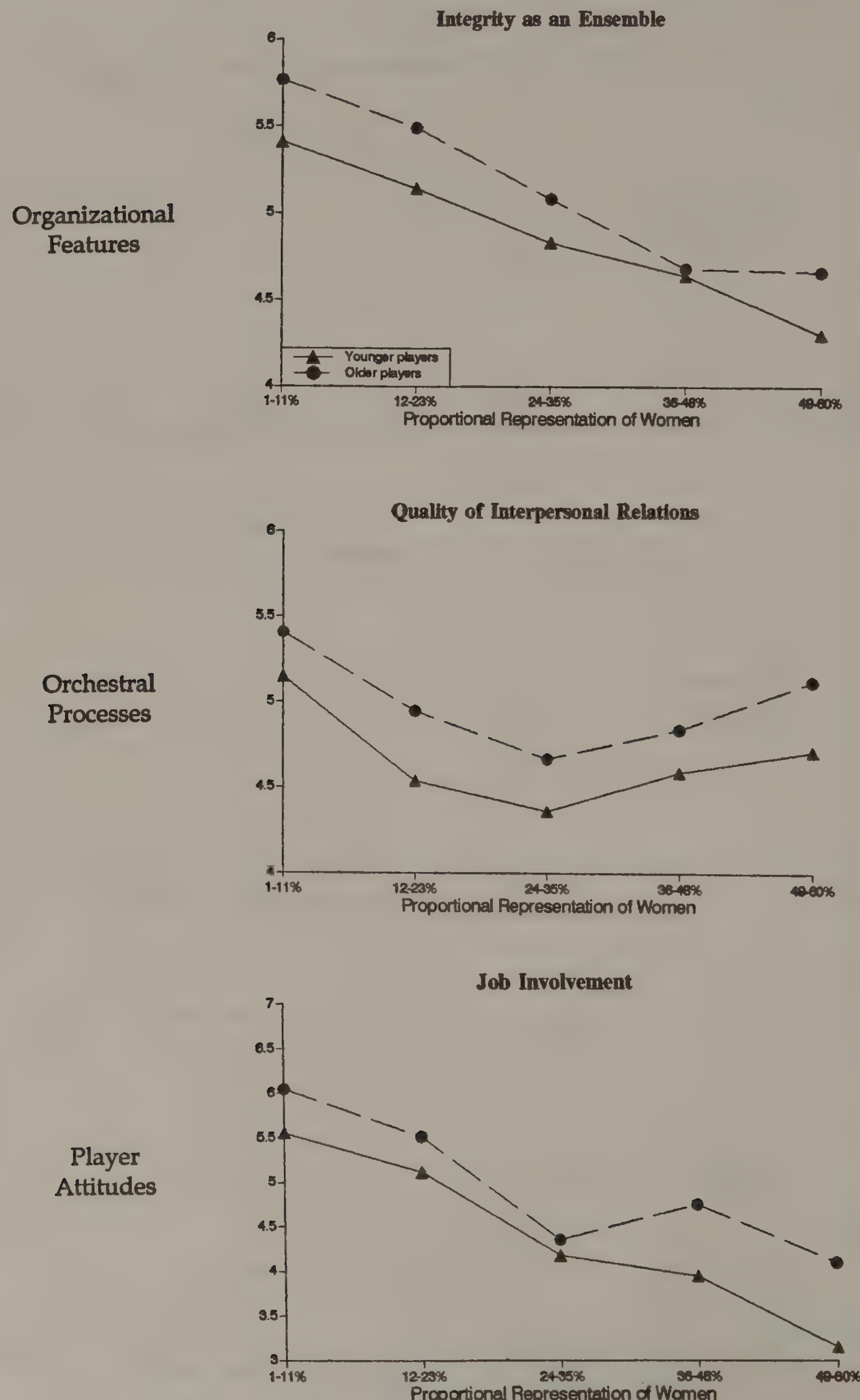
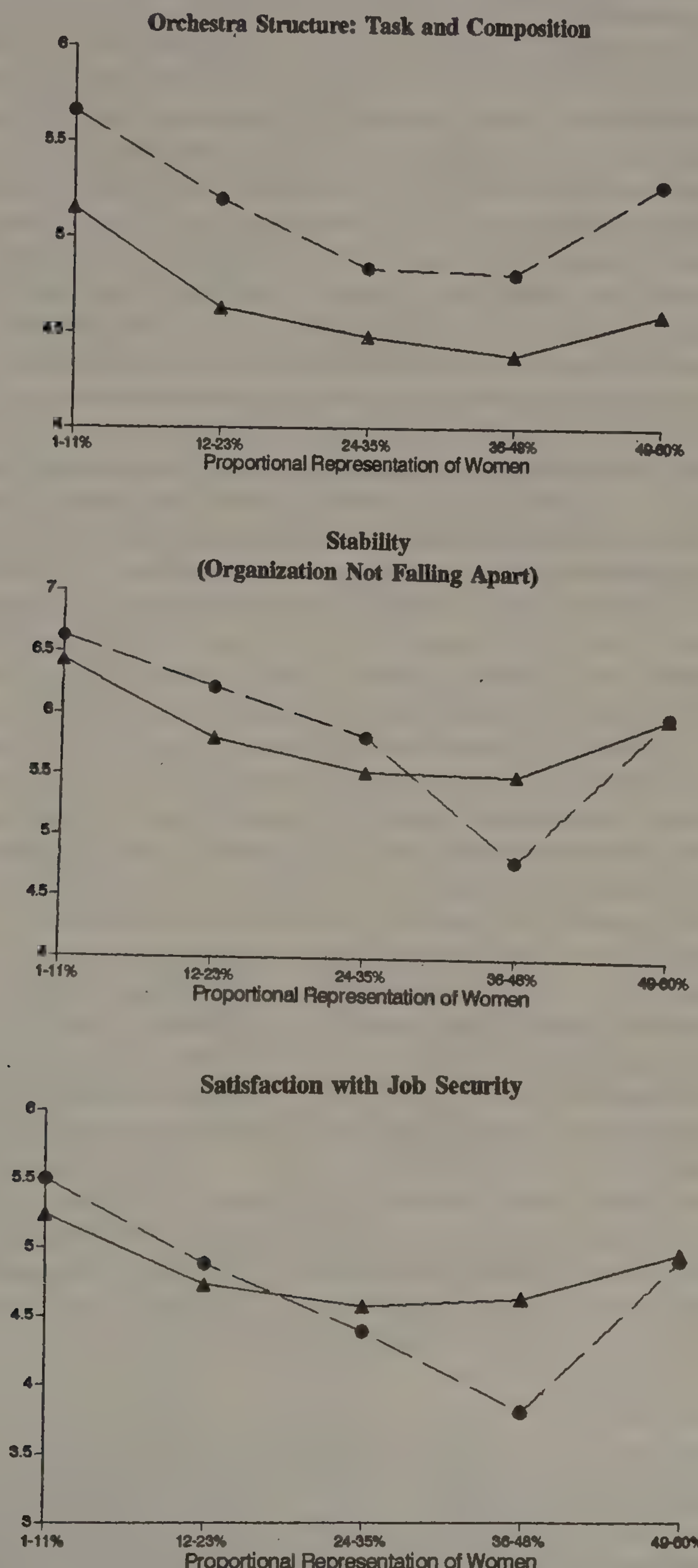


FIGURE 3: Proportion Effects for Younger vs. Older Players (Continued)

Organizational Features

Orchestral Processes

Player Attitudes



both age groups the previously reported results according to the proportion of women. The gender composition main effect is not an artifact of player age and experience.

Instruments. Another rival explanation derives from the sex-role typing of the instruments of the orchestra (Griswold & Chroback 1981). Women are significantly overrepresented in the first and second violins, violas, and cellos; they are underrepresented in the woodwind, percussion, bass, and brass sections (Allmendinger & Hackman 1993a). Thus, women play mainly in the largest sections and are not as prominent in orchestras as they would be if they were playing instruments for which there are fewer players per part.[7] This raises the possibility that the main effect for proportion of women is really an instrument effect — i.e., that it is due to differences between the violins, violas, and cellos on the one hand (which we will call female type instruments) and the basses, woodwinds, brass, and percussion on the other (which we will call male type instruments). Therefore, we examined the relationship between the proportion of women and our six representative measures separately for individuals who play female-type and male-type instruments. The findings, shown in Figure 4, once again replicate the main effect findings for gender composition. The proportion effect is not an artifact of instrument type.

Standing. When women enter a profession previously dominated by males, they usually appear initially in lower status organizations. Lie and O'Leary (1990) show this to be true for the academic profession; we find it as well for symphony orchestras. In each of the four countries of our sample (and especially in the U.S.), there is a significantly higher proportion of women in smaller, less affluent, and less prestigious minor orchestras than there is in relatively larger, richer, and more prestigious major orchestras. (The proportions of women in minor and major orchestras in each country are, respectively: West Germany, 20% versus 13%; East Germany, 22% versus 13%; United Kingdom, 38% versus 26%, and U.S., 52% versus 24%.) It is possible, therefore, that obtained main effects for gender composition are actually reflecting orchestra standing — specifically, that minor orchestras have both a higher proportion of women and lower scores on the dependent measures than do major orchestras. To test this possibility, we examined the relationship between proportion of women and our representative measures separately for major and minor orchestras. As seen in Figure 5, *major* orchestras show the more pronounced drop on our dependent measures; it is the relatively less well-off minor orchestras that are mostly responsible for the upturn that occurs for many variables when the proportion of women approaches 50%. Although unanticipated, this finding is precisely what would be expected from the research literature on gender segregation (e.g., Bielby & Baron 1986; Reskin & Roos 1990).

Summary

The findings reported in Table 3 — namely, that almost all measures of perceived organizational features, orchestral processes, and player motivation and satisfaction show an initial decline as the proportion of women in an orchestra increases — are strong and apparently stable. For many measures, this downward trend flattens or reverses when the proportion of women approaches

50%, yielding curves that are remarkably similar to those reported by Pfeffer and Davis-Blake (1987) for the relationship between organizational gender composition and salaries (compare our Figure 2 to their Figures 1 and 2, p. 15). This upturn appears to be driven mainly by the reports of players in *minor* orchestras, about which we will have more to say later.

All our dependent variables are based on self-reports by orchestra players. That presents no interpretative problem for the measures of player motivation, satisfaction, and relationship quality; for those measures, it is the players' own states and experiences that they are reporting. For the measures of organizational features, however, interpretations must be more cautious, since players' affective states may distort their perceptions of the properties of their organizations. It is instructive, therefore, to examine the relationship between gender composition and the organizational features of orchestras using data that are independent of players' perceptions. Project researchers independently assessed several organizational properties while on site at each orchestra, using a structured observational guide. The means of the researchers' assessments of two organizational features — integrity as an ensemble and orchestra structure — were computed for orchestras in the top, middle, and bottom thirds of the gender composition distribution. For integrity as an ensemble, the means (from lowest to highest proportion of women) are 5.27, 5.06, and 4.00 ($F = 9.74$, $p < .01$); for orchestra structure, the means are 5.27, 5.11, and 4.75 ($F = 1.8$, n.s.). This pattern matches that obtained using members' survey descriptions of the features of their orchestras.

We turn next to results for our second and third research questions, which explore the interactions between orchestra gender composition and (1) the gender of individual respondents, and (2) orchestras' national contexts. These findings should help identify and illuminate the mechanisms that give rise to the main effects reported above.

COMPOSITION-GENDER INTERACTIONS

The main effects reported thus far describe the combined responses of female and male players. Yet organizational gender composition may shape experiences, perceptions, and attitudes of the new entrants, women, more strongly than those of the old hands, men — or vice versa. This possibility can be tested by examining the interaction between respondent gender and orchestra gender composition. If the gender composition does affect women more than men, then (1) there should be significant gender-by-composition interactions for many of our dependent measures, and (2) the slope of the means across the gender composition spectrum should be greater for women than for men.

Table 4 shows that the first condition is met but that the second is not. Ten significant gender-by-composition interactions were obtained, but for all of them it is *men* who respond more strongly to changes in organizational gender composition. Moreover, the standing of the two genders on all dependent measures reverses between token and relatively balanced gender composition; men's reports are more favorable when there are few female players, whereas women's reports are more favorable when there are many. The typical pattern

FIGURE 4: Proportion Effects for "Male" vs. "Female" Instruments

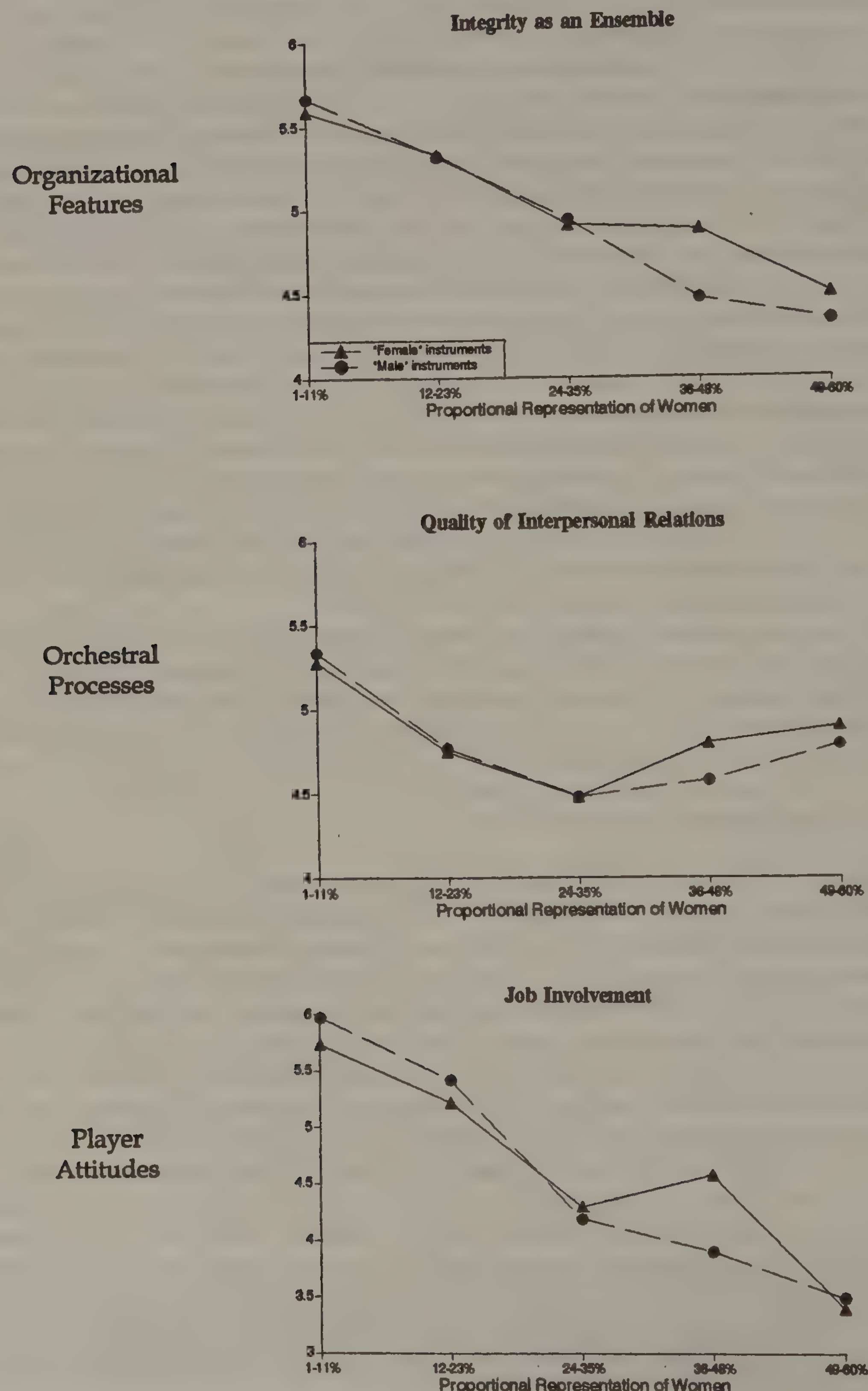


FIGURE 4: Proportion Effects for "Male" vs. "Female" Instruments

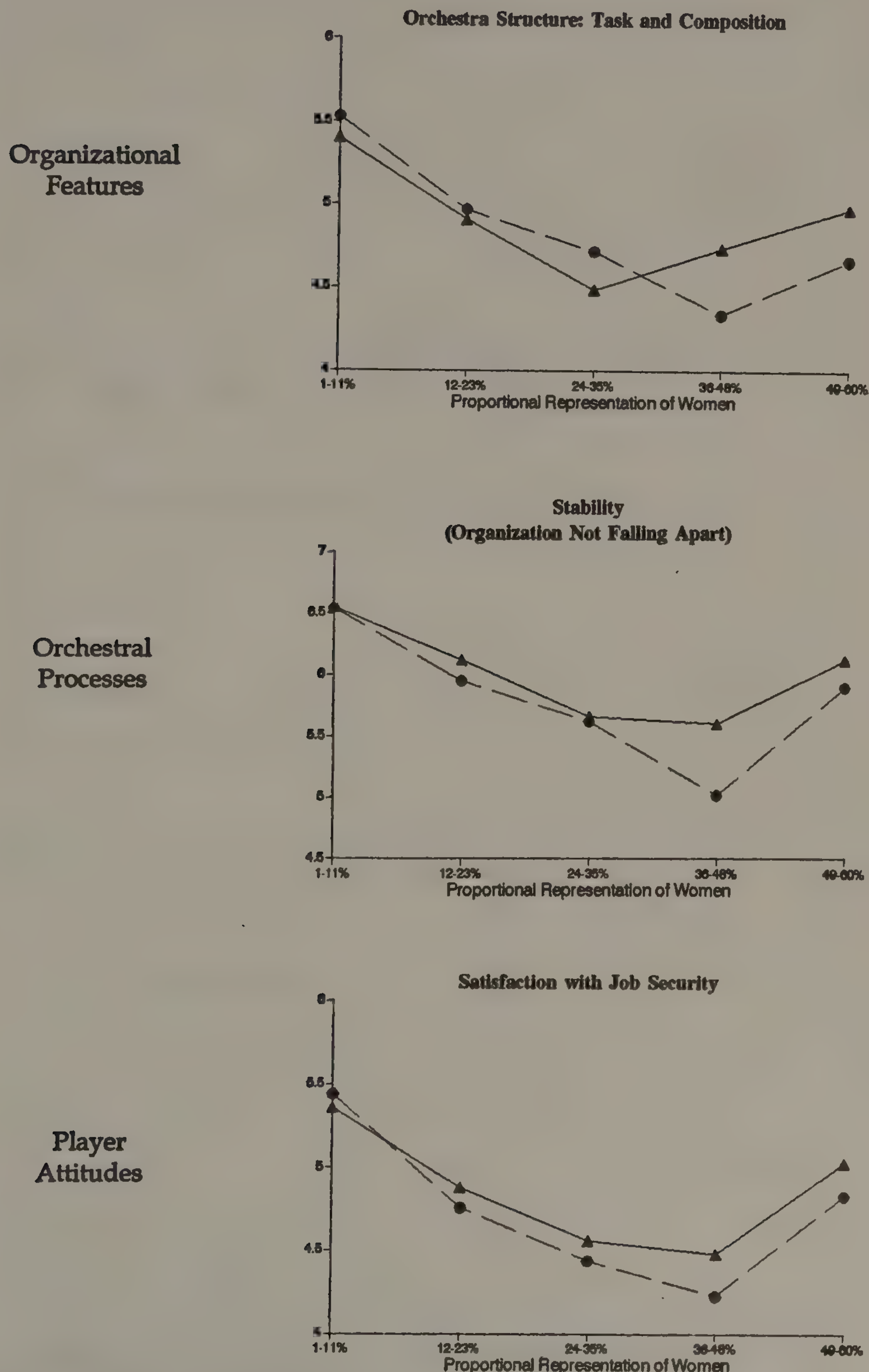


FIGURE 5: Proportion Effects for Major vs. Minor Orchestras

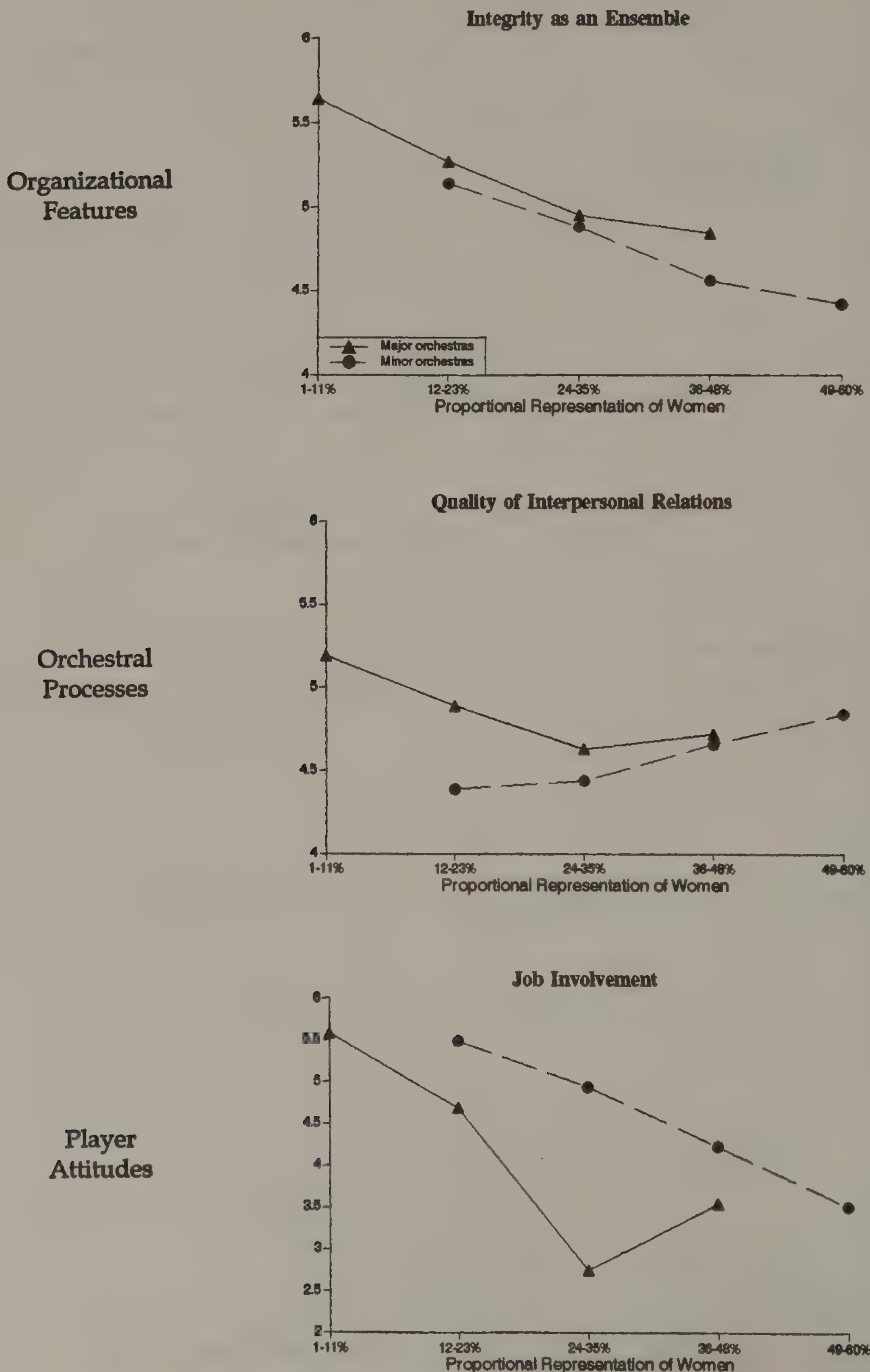
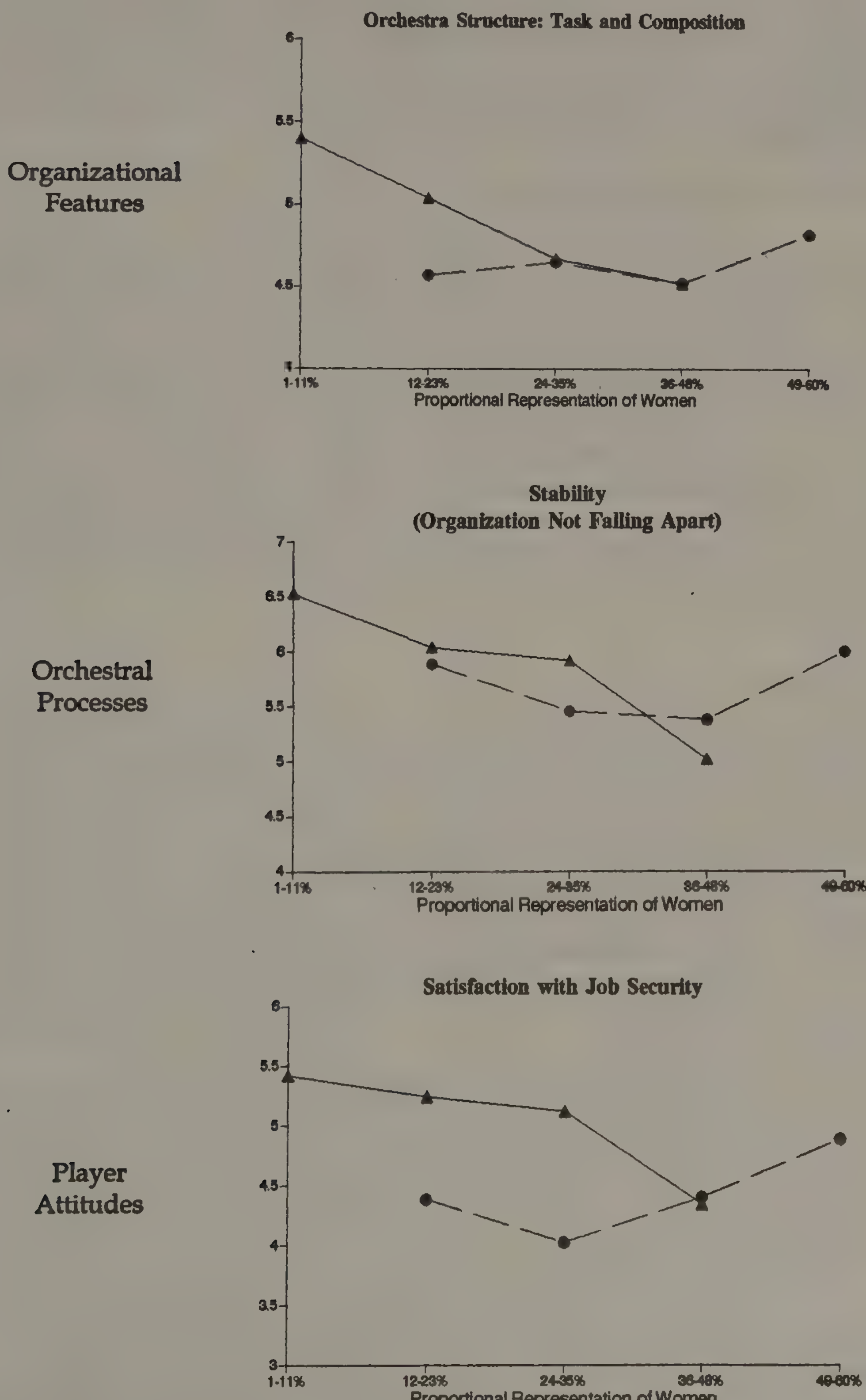


FIGURE 5: Proportion Effects for Major vs. Minor Orchestras (Continued)



of the significant interactions is shown graphically for one variable, integrity as an ensemble, in Figure 6.[8]

It is noteworthy that for most of the dependent measures in Table 4, *both* men and women have higher scores at lower levels of female representation than at higher levels. This pattern, which is consistent with the main effect findings reported earlier, also fits with those of Wharton and Baron (1987, 1991), who reported that women as well as men are more satisfied with their jobs in male-dominated than in mixed work settings.

COMPOSITION-COUNTRY INTERACTIONS

Significant interactions between gender composition and country were obtained for 16 of the 22 dependent measures (see Table 2). Although effect sizes for these interactions are smaller than those of their component variables, they do suggest that the impact of changes in orchestra gender composition depend significantly on national context. Unfortunately, country-by-composition interactions necessarily involve twenty organization-level means (four countries by five proportions); given that our total sample of orchestras is 72, many cells that would be of great interest have only one or two entries — and a few have none at all. Therefore, we must rely mainly on archival and qualitative data to explore the mechanisms by which national contexts affect the dynamics and outcomes of the entry of women into symphony orchestras. We consider each country in turn.

United States

Orchestras in the U.S. more closely approach relatively balanced gender composition than do those in the other three nations. This country has had by far the most aggressive social policies of the four nations regarding equal employment opportunities. In addition, ICSOM (the union-related organization that serves players in major orchestras) some years ago mounted a campaign to improve the fairness of orchestra recruitment, audition, and selection processes. Union orchestras began to publicly advertise all open positions, to hold "blind" auditions (in which qualified applicants play behind a screen), and to significantly constrain the right of music directors to make employment decisions autonomously. The result has been a significant increase in the number of women in U.S. orchestras.

Minor orchestras in the U.S. are particularly informative because they are the ones for which gender composition is most balanced of all. Indeed, almost all the orchestras whose female membership exceeds 50% are minor orchestras located in U.S. Women in these orchestras have achieved a level of legitimacy and acceptance that, so far, is uncommon either in major U.S. orchestras or in orchestras in other countries.

Moreover, it is U.S. minor orchestras that are mainly responsible for the upturn at the end of the monotonic decline exhibited by many measures as the proportion of women increased. It is noteworthy that these improvements occurred despite the resource constraints that challenge virtually all minor orchestras. Although the dynamics of gender integration are likely to differ substantially for major and minor orchestras, it is possible nonetheless that

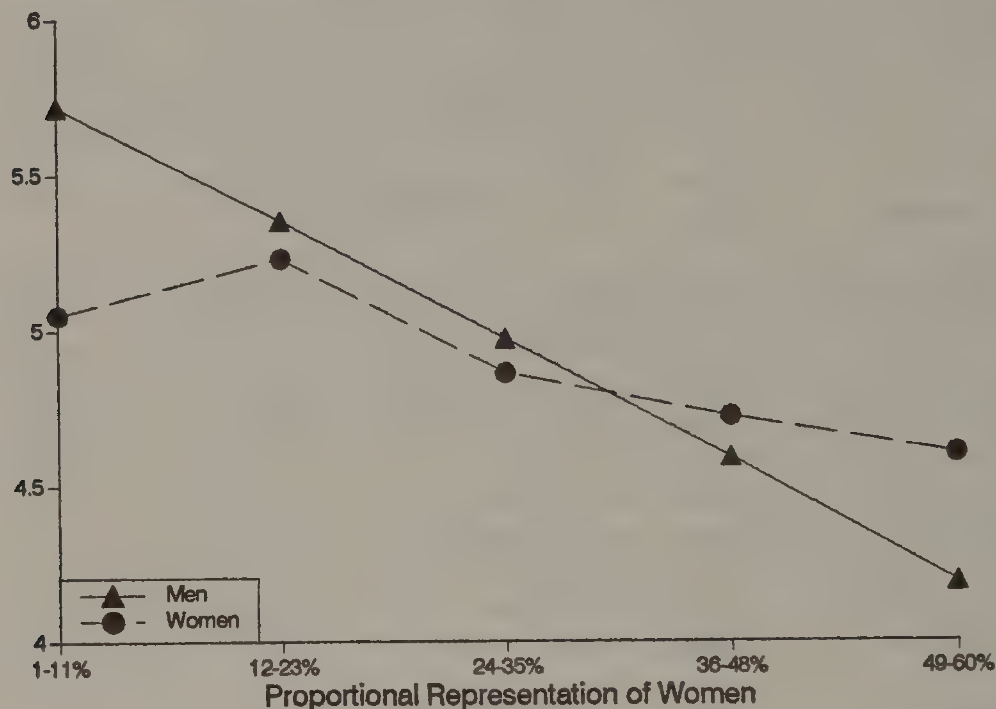
TABLE 4: Significant Interactions between Individual Gender and Orchestra Gender Composition[a]

		Proportion of Women (%)					F ratio (Gender by Proportion [Interaction])
		0-11	12-23	24-35	36-48	49-60	
Integrity as an ensemble	Men	5.72	5.35	4.97	4.59	4.19	5.0**
	Women	5.05	5.23	4.86	4.72	4.60	
Orchestra structure	Men	5.50	4.96	4.68	4.40	4.63	4.8**
	Women	4.91	4.86	4.52	4.67	5.01	
Player involvement	Men	4.64	4.26	4.04	3.59	3.67	4.1*
	Women	4.18	4.16	3.89	4.07	4.02	
Recognition for good work	Men	6.36	5.69	5.41	4.84	4.22	4.2*
	Women	6.05	5.85	5.12	5.33	5.08	
Fair/effective recruitment	Men	6.04	5.58	5.18	4.73	4.51	3.1*
	Women	5.40	5.33	5.22	5.25	5.08	
Music director coaching	Men	2.94	2.79	2.69	2.73	2.72	3.3*
	Women	2.48	2.81	2.87	3.04	3.14	
Relationship quality	Men	5.41	4.80	4.54	4.53	4.58	6.4**
	Women	4.66	4.63	4.41	4.88	5.04	
Musical outcomes	Men	4.86	4.63	4.87	5.21	4.94	4.8**
	Women	4.28	4.34	5.03	5.39	5.50	
Organizational stability	Men	6.59	6.00	5.62	4.84	5.56	4.2*
	Women	6.34	6.09	5.68	5.95	6.37	
Growth satisfaction	Men	5.27	4.98	4.63	4.63	4.96	3.5*
	Women	4.82	4.56	4.62	5.02	4.99	

[a] Means are based on weighted individual-level data from 979 respondents.

* $p < .01$ ** $p < .001$

FIGURE 6: Gender-by-Composition Interaction for "Integrity as an Ensemble"



minor orchestras in the U.S. are providing an advance look at what eventually could become commonplace as female players achieve full legitimacy in symphony orchestras around the world.

United Kingdom

Although women are relatively well represented in British orchestras, their participation in the national labor force is considerably lower than in the U.S. Symphony orchestras in this country, then, are in the vanguard of the movement of British women into the work force and the professions.

The United Kingdom was unique among the nations studied in showing modest *improvement* in some scores on organizational features as the proportion of women increased. It may be that orchestra members, when comparing themselves to other British organizations, view themselves and their orchestras as doing rather well. And, one could argue, the larger the proportion of women, the more likely it is that orchestra members will have such perceptions.

A second explanation for the atypical relationship between proportion of women and certain of our measures derives from the player selection processes used by British orchestras. There are three relatively distinct groups of orchestras in the United Kingdom: (1) four major orchestras located in London, each a worker cooperative in which players are compensated on a "per service" basis; (2) six regional orchestras, not cooperatives, whose players hold annual

contracts, and (3) four British Broadcasting Corporation orchestras, all of whose members are civil servants. Because British orchestras generally do not use blind auditions, members know exactly who they are hiring and, especially in the London cooperatives and in the regional orchestras, pride themselves on selecting players who "fit in" with the orchestra — socially as well as musically. Moreover, since neither the government nor the musicians' union forces the hands of orchestras in hiring decisions, those women who are chosen for membership are likely both to be welcomed by the players who hired them and to have work styles that are generally congruent with the existing and predominantly male organizational culture.

Women in British orchestras, therefore, may be more inclined than their counterparts in other countries to perpetuate the existing orchestral culture — both in their behavior in the orchestra and when they participate in the selection of new players — than would women who had been selected solely on the basis of technical prowess. Women in British orchestras may not be adding as much diversity to their ensembles as would be the case if blind auditions were used. On the other hand, relations between male and female players almost certainly are more collegial than they otherwise would be — attested to by the fact that players in U.K. orchestras report higher satisfaction with work relationships than do players in any of the other countries in our sample (Allmendinger & Hackman 1992).

The future pattern of gender dynamics in British orchestras is far from certain. Significant tensions may develop as women approach majority status and men have to confront the fact that women's gains have occurred, to some extent, at their expense. Yet if the national as well as the orchestral work force eventually approaches gender balance, then player attitudes and relationships in these orchestras may look very much like what is now characteristic of many minor orchestras in the U.S.

East and West Germany

The situation in the former East Germany is almost the opposite of that in the United Kingdom. In the late 1980s, East German women had a higher rate of participation in the national work force, by a wide margin, than did women in any other country we studied — but symphony orchestras were still as they historically had always been, predominantly male. This fact highlights an exception to the general policy of "democratizing" the work force espoused by the leaders of the socialist state. Although women in the rest of East German society were given incentives to enter the work force and organizations were rewarded for employing them, this was deliberately not done for orchestras. Indeed, liberal government policies regarding leaves for childbirth and child care provided *dis*incentives for orchestras to hire women since substitute players were a scarce commodity. In East Germany, symphony orchestras became islands of male dominance in a sea of relative gender balance.

In West Germany, symphony musicians are civil servants paid with public funds. Although public policy dictates gender-blind employment opportunity, that policy has not been enforced for symphony orchestras. Moreover, the union that represents orchestra musicians has not actively promoted the entry of

women into orchestras. The result is that West German women have low rates of participation in *both* the national work force and in symphony orchestras.

These differences between the two Germanys allow empirical exploration of national-level influences on orchestral gender dynamics. In the German version of our survey, we included several items that asked specifically about the quality of gender relationships in respondents' orchestras (for details, see Allmendinger & Hackman 1993b). Findings for two of those items are informative. East German musicians agreed more strongly with the item, "In this orchestra, men and women support each other and work together toward common goals," whereas West German musicians were higher on the item, "Men and women are treated differently in this orchestra." Thus, even though the proportions of women in orchestras in the two German nations were almost identical, the gender *climate* was more favorable in East German orchestras, where women had a high rate of participation in the national work force, than in West German orchestras, where they did not. As also was found for the U.S. and the United Kingdom, we see that the national context does indeed affect how individuals respond to their local gender circumstances.

Discussion

There are large differences in the effect sizes of individual gender, organizational gender composition, and national context on the perceptions, attitudes, and motivation of orchestra players. Overall, respondent gender controlled less than 1% of the dependent variable variance. Members of a given orchestra in a given country, women and men alike, have attitudes that are far more similar to one another than they are to the attitudes of members of other orchestras. A symphony orchestra is a "strong situation" (Weiss & Adler 1984), in which institutional properties such as task demands and social norms are clear and powerful. These situational forces, apparently, dominate the influence of gender on members' affective reactions to their work.

This conclusion is reinforced by the findings for organizational gender composition (i.e., the proportion of members who are female) and for national context (i.e., differences among countries in their receptivity to women's participation in the national labor force). For our measures of member perceptions and attitudes, it is these structural and contextual variables that make the most difference.

ORGANIZATIONAL GENDER COMPOSITION

The traditional dictum, "the more, the better" as it applies to women entering traditionally male organizations is too simple. Our findings from symphony orchestras suggest that the problems encountered during the early stages of gender integration are unlikely to be resolved by simply increasing the proportion of women beyond token levels, a conclusion also reached by Wharton and Baron (1987:586). Although our findings are based on cross-sectional data, the interactions obtained between orchestra gender composition and both individual gender and national context make it possible to draw plausible conclusions about the dynamics of gender recomposition processes.

Specifically, our data suggest that gender dynamics differ qualitatively at three succeeding stages: (1) when an organization has but a token number of women (e.g., 10% or fewer female members), (2) when it is in transition (e.g., around 10-40% female), and (3) when it has become relatively gender balanced (e.g., 40-60% female).[9]

Organizations with Token Female Representation

Following Kanter's (1977a) pioneering study, it was generally accepted that the work life of male members of highly skewed organizations usually is quite satisfactory — but that the situation for their female colleagues is dissatisfying, stressful, and alienating. Our findings suggest that the picture may be more complex than that. Even though women are less satisfied with their orchestras and their work lives at token levels of representation, they are nonetheless more satisfied then than they are at the next higher levels of female participation.

It is no doubt true that, in male-dominated organizations, neither the organization nor its members are obtaining the benefits (such as personal learning and improved task performance) that compositional diversity can bring. Indeed, our qualitative data suggest that many women find that there are strong incentives for them to keep a low profile, to behave closely in accord with existing orchestral norms, and generally to be as nonintrusive as possible. This stance is costly to the orchestra because it protects majority members from exposure to unfamiliar perspectives and from the need to scrutinize and reconsider traditional behavioral norms. On the other hand, it increases the likelihood that the orchestra will run smoothly and it minimizes the kinds of tensions that characteristically develop when an orchestra moves into what we are calling the transitional stage of gender composition.

Organizations in Transition

Our findings show that organizational life for both female and male members takes a qualitative turn for the worse when the proportion of women reaches the transitional stage (i.e., 10-40%). For the first time, there are enough women to form a significant subgroup in the orchestra. Women now have a set of colleagues with whom they can share and test their perceptions, thoughts, and feelings (Zanna, Crosby & Lowenstein 1987). Moreover, it is now feasible for them to become a political force in the orchestra — something that is possible for a group, but rarely for an individual, in an established organization. And, as women become both more numerous and more powerful, men can no longer ignore the threat women pose to their occupational status and organizational control.

Together, these processes result in tightened identity group boundaries for both genders, increased cross-group stereotyping and conflict (Wharton & Baron 1987), less social support across gender boundaries (South et al. 1982), and heightened personal tension for everyone. These problems do not automatically correct themselves as the proportion of women grows; indeed, there are empirical as well as theoretical reasons to expect them to worsen as an organization moves through the transitional phase of gender integration (e.g., Blalock 1967; Taylor et al. 1978; Tsui, Egan & O'Reilly 1992). In our study, it was

not until women achieved a representation of over 40% that robust signs of improvement — the "tipping point" identified by Pfeffer and Davis-Blake (1987) — were first seen.

Relatively Balanced Organizations

In these organizations, all members have plenty of people like themselves with whom to do social comparisons and from whom to seek support. Moreover, members of both gender groups are likely to feel fully legitimate in the organization — neither closely scrutinized nor especially threatened. Under such circumstances, relationships between men and women should begin to stabilize. And with both constituent groups strong, it now becomes feasible to develop intergroup relations that are characterized more by interdependence and mutual support than by conflict (Smith 1982).

Our conclusion that such changes are not likely to be seen until after the proportion of female members exceeds 40% is identical to that reached by Martin (1985:329). And, even at that point, the monotonic decline seen for almost all our dependent variables continued unabated for four of them: three measures of members' perceptions of organizational features (the degree to which the orchestra has integrity as an ensemble, the degree to which players are recognized for good work, and the fairness and effectiveness of recruiting and auditioning practices) and the measure of individual job involvement. It remains an open question whether, at even higher levels of female representation, these measures also would have shown a reversal — and whether those measures that did reverse eventually would have surpassed the levels found in orchestras that had but token female representation.

CONCLUSION

To conclude, we step back and place our findings in larger context. What are the implications of the patterns we have reported for understanding, influencing, and researching the entry of women into organizations that have long been dominated by men? We propose three explanations:

(1) This is how the world works. Things really do get more difficult — for women, for men, and for the social system — as more women enter traditionally male organizations, at least until a state of gender balance is achieved. This generalization applies to all types of organizations and probably to all western cultures. Behavioral science interventions may be of some use in helping members of organizations undergoing demographic transition learn how to deal constructively with the increasing diversity that they are experiencing — but they are in opposition to natural social system forces.

(2) Symphony orchestras are unique. The type of organization we studied, the professional symphony orchestra, is unique in a number of ways. The work is to perform music. Accomplishing that work requires extraordinary levels of member coordination. Playing in a symphony orchestra is a high status profession, one with high barriers to entry. An orchestra is a true ensemble, but members work under the strong direction of a leader who is almost always male. And, as an institution, the symphony orchestra has a tradition of male exclusivity that is both longstanding and deeply rooted. It may be that

organizational gender composition would control less variation in attitudes in other types of organizations, such as schools or certain businesses, whose traditions are less strongly male and whose charters are not so unique. Moreover, the decline we found for many measures as the proportion of women increased might have been less steep, or might even have been reversed, in organizations with different gender traditions.

(3) A global transition in gender relations is occurring. There is a world of difference between a minor orchestra in the U.S. whose membership is 60% female and a major West German symphony that has but two or three female players. The former orchestra has completed its transition from male exclusivity to gender balance; the latter has still to begin. What we have in our findings, then, is but a snapshot of a moving phenomenon — and that phenomenon is proceeding at different rates for different types of organizations in different countries. Researchers can study the change process, interventionists can try to help it along, and managers can behave in ways that promote or impede it. But, in this view, gender recomposition in the workplace will inexorably move forward, regardless of what these and other professionals do.

The gender recomposition process is still underway in orchestras around the world and is at a relatively early stage in many of them. For this reason, it is not possible to rule out or to unambiguously affirm any of the three alternative explanations. Overall, we find the third alternative the most plausible and the most consistent with our evidence — but it will be necessary to wait for longitudinal data to become available before it can be definitively tested. Regardless of which interpretation one prefers, the present data do make clear that research done solely at any single level of analysis (be it the individual level, the organizational level, or the national level) and in any single nation is unlikely, by itself, to adequately explain and inform organizational gender dynamics. For it is in the *interactions* among personal, organizational, and contextual factors that the patterns of gender dynamics explored in this article come to life and play out their consequences.

Notes

1. Although Kanter's theory deals with men's and women's relative power and career opportunities as well as with their numbers, most studies prompted by her research have focused on the attitudes and experiences of token members (i.e., those whose subgroup consists of 15% or less of the total membership). Some subsequent findings support her propositions (e.g., Spangler, Gordon & Pipkin 1978; Yoder, Adams & Prince 1983); others do not (e.g., Fairhurst & Snavely 1983; South et al. 1982). Partly in response to these mixed empirical results, several alternative models of gender composition phenomena have been proposed in recent years. These models give relatively less emphasis to numbers and more to factors such as gender roles (Eagly 1987), expectation states (Ridgeway & Diekema 1992), and the structural features of social systems (Martin 1985).

2. It may be difficult for men to report, or even to explicitly acknowledge, such threats. Instead, as noted by one of the reviewers of this article, they may be more comfortable reporting that the organization does not function as well, or that it has a poorer climate, than publicly acknowledging that they are experiencing a threat to their status or control.

3. The power of the canon was especially evident in East Germany. Official policy during the socialist regime was for orchestras to emphasize in their programs the music of contemporary composers from socialist countries. To assess the impact of this policy, we selected one regular

concert program from each orchestra in our sample and noted the birth dates of the composers whose works were played. The median composer birth date for East Germany was 1809, compared to 1857 for West Germany, 1862 for the U.S., and 1864 for the United Kingdom. Despite official policy, East German orchestras continued to perform the standard repertoire — indeed, exhibiting greater commitment to the classical tradition than did orchestras in more market-oriented countries.

4. Kanter uses six categories to describe the proportional representation of women in organizations: uniform groups (no women), skewed groups (1-11% women), low-tilted groups (12-23%); high-tilted groups (24-35%), low-balanced groups (36-47%), and high-balanced groups (48-60%). For comparability with the numerous studies that use these categories, we also employ them in most of our analyses.

5. Assistance in enumerating the population of orchestras and in gaining access to them was provided by the American Federation of Musicians (AFM), the American Symphony Orchestra League (ASOL), the Arts Council of Great Britain (ACGB), the European Conference of Symphony Orchestras (ECSO), the International Conference of Symphony and Opera Musicians (ICSOM), and the Regional Orchestra Players Association (ROPA). We give special thanks to the following individuals: Kenneth Baird (ACGB), Brad Buckley (ICSOM), Rosemary Estes (ROPA), Catherine French (ASOL), Ken Haas (Boston Symphony Orchestra), and Lew Waldeck (AFM).

6. The three rival explanations are addressed serially because the analytic model required to assess them simultaneously generates an unacceptably imbalanced design (e.g., it generates a number of entirely empty cells). Nonetheless, an omnibus test was run to provide a rough assessment of the robustness of the main-effect findings for orchestra gender composition (Table 3). A hierarchical general linear model was constructed with gender composition entered as the penultimate term (i.e., after individual gender, orchestra standing [major versus minor], nation, age/experience, and sex-type of instrument [discussed later]; and before specific orchestra, treated as a random effect). This analysis yielded statistically significant gender composition effects for 18 of our 22 dependent variables (compared to significant effects for 19 of those variables in the main-effect analysis reported in Table 3). These findings affirm that significant gender composition effects remain even after those associated with other individual, organizational, and contextual factors are controlled. An additional omnibus analysis was run to assess the statistical reliability of the quadratic component of the main effect findings (i.e., the upturn found for some variables at the high end of the gender composition continuum). All quadratic effects found to be significant in the main analyses (reported in the rightmost column of Table 3) remain so after the control variables are introduced.

7. The composition of a full-size symphony orchestra is roughly as follows: 20 first violins, 20 second violins, 12 violas, 12 cellos, 10 basses, 16 woodwinds, 17 brass, and 4 percussion and other (Previn 1979, chap. 1). Gender differences across sections are substantial. In an analysis of the relative potency of section, orchestra standing (major versus minor), and country on the representation of women, section controlled 29% of the variance, compared to 14% and 6% for country and standing respectively.

8. This pattern does not hold for the two variables in Table 4 that have directly to do with musical matters — music director coaching and perceived musical outcomes. Women's reports for these two measures become increasingly favorable the larger the proportion of women in the orchestra, whereas men's reports exhibit no consistent pattern.

9. We do not address here either homogeneously male or dominantly female organizations because no such orchestras were in our sample. Reports about the Vienna Philharmonic Orchestra, one of the few remaining all-male professional symphony orchestras in the world, are that it is an orchestra whose members work together especially well as an ensemble (Blaukopf & Blaukopf 1986). It would be informative to compare the structure and functioning of that orchestra to an all-female ensemble.

APPENDIX

Measures Derived from the Player Survey

1. Organizational features of the orchestra

Integrity as an ensemble. The degree to which the orchestra operates as an intact, relatively stable ensemble, with clear and well-accepted norms of behavior. (8 items)

Orchestra structure: task and composition. The degree to which the work of the orchestra is well designed (i.e., the orchestra's tasks are challenging and rich in feedback) and the orchestra itself is well composed (i.e., players have a sufficient level and mix of talent and experience for their collective work). (5 items)

Player involvement and opportunities. The degree to which players have information and influence in the orchestra, share in its leadership, and have opportunities to improve their position within it. (8 items)

Adequacy of resources. The adequacy of the orchestra's facilities and material resources. (2 items)

Single items retained:

Recognition. "Excellent playing pays off in this orchestra."

Recruitment. "The orchestra's recruitment and audition process is fair and effective."

2. Behavior of the music director

Direction-setting. The degree to which the music director sets a clear and engaging artistic direction for the orchestra. (5 items)

Organizational initiatives. The degree to which the music director takes organizational initiatives (both internally, such as by working with management to improve the organizational structure or the recruitment process, and externally, such as by building community support or negotiating recording contracts). (7 items)

Coaching. The degree to which the music director makes himself or herself available to orchestra members and actively coaches individual players and sections. (5 items)

Single item retained:

Music director authority. "In this orchestra, the music director is the only *real* boss."

3. Orchestral processes

Quality of relationships. The quality of the working relationships among orchestra members. (7 items)

Musical outcomes. The degree to which the orchestra is improving (rather than slipping) musically. (3 items)

Single items retained:

Finances. "Our orchestra is getting stronger and stronger financially."

Stability. "As an organization, our orchestra shows signs of falling apart." (reverse scored)

4. Player motivation and satisfaction

General satisfaction. The respondent's overall satisfaction with his or her job and work. (3 items)

Internal work motivation. The degree to which the respondent is self-motivated to perform well (i.e., feeling good when having performed well, and feeling bad when having performed poorly). (3 items)

Job involvement. A single item: "I live, eat, and breathe my job."

Satisfaction with specific organizational features:
Compensation (2 items); Job security (2 items); Management (3 items); Work relationships (3 items); Growth opportunities (4 items)

References

Alexander, Victoria D., and Peggy A. Thoits. 1985. "Token Achievement: An Examination of Proportional Representation and Performance Outcomes." *Social Forces* 64:332-40.

Allmendinger, Jutta, and J. Richard Hackman. 1992. *Bringing Music to People: Continuity and Discontinuity in East German Orchestras.* Report no. 3, Cross-National Study of Symphony Orchestras, Harvard University.

____. 1993a. *The More the Better? On the Inclusion of Women in Professional Symphony Orchestras.* Report no. 5, Cross-National Study of Symphony Orchestras, Harvard University.

____. 1993b. *Männer und Frauen in Organisationen: Das Beispiel west- und ostdeutscher Orchester.* Institut für Soziologie, Universität München.

Allmendinger, Jutta, J. Richard Hackman, Larissa Kowal-Wolk, and Erin V. Lehman. 1993. *Methods and Measures for the Cross-National Study of Symphony Orchestras.* Report no. 4, Cross-National Study of Symphony Orchestras, Harvard University.

Bielby, William T., and James N. Baron. 1986. "Men and Women at Work: Sex Segregation and Statistical Discrimination." *American Journal of Sociology* 91:759-99.

Blalock, Hubert M., Jr. 1967. *Toward a Theory of Minority Group Relations.* Wiley.

Blaukopf, Herta, and Kurt Blaukopf. 1986. *Die Wiener Philharmoniker.* Zsolnay.

Block, Adrienne F., and Carol Neuls-Bates. 1979. "Historical Introduction." *Women in American Music: A Bibliography of Music and Literature.* Greenwood Press.

Blum, Linda, and Vicki Smith. 1988. "Women's Mobility in the Corporation: A Critique of the Politics of Optimism." *Signs* 13:528-45.

Cockburn, Cynthia. 1985. *Machinery of Dominance: Women, Men and Technical Know-how.* Pluto Press.

Dion, Kenneth L. 1985. "Sex, Gender, and Groups: Selected Issues." Pp. 293-347 in *Women, Gender, and Social Psychology,* edited by Virginia E. O'Leary, Rhoda K. Unger, and Barbara S. Wallston. Erlbaum.

Eagly, Alice H. 1987. *Sex Differences in Social Behavior: A Social-Role Interpretation.* Erlbaum.

Edwards, J. Michele. 1991, July. *Transformation and the American Orchestra: Women and Gender Issues.* Paper presented at a conference on women and gender, London.

Fairhurst, Gail T., and B. Kay Snavely. 1983. "A Test of the Social Isolation of Male Tokens." *Academy of Management Review* 26:353-61.

Faludi, Susan. 1991. *Backlash: The Undeclared War against American Women.* Crown.

Frisbie, W. Parker, and Lisa Neidert. 1976. "Inequality and the Relative Size of Minority Populations: A Comparative Analysis." *American Journal of Sociology* 82:1007-30.

Griswold, Philip A., and Denise A. Chroback. 1981. "Sex-Role Associations of Music Instruments and Occupations by Gender and Major." *Journal of Research in Music Education* 29:57-62.

Hackman, J. Richard. 1992. "Group Influences on Individuals in Organizations." Pp. 199-267 in *Handbook of Industrial and Organizational Psychology.* Vol. 3, edited by Marvin D. Dunnette and Laetta M. Hough. Consulting Psychologists Press.

Hackman, J. Richard, and Greg R. Oldham. 1976. "Motivation through the Design of Work: Test of a Theory." *Organizational Behavior and Human Performance* 16:250-79.

Hoffman, L. Richard. 1965. "Group Problem Solving." Pp. 99-132 in *Advances in Experimental Social Psychology.* Vol. 2, edited by Leonard Berkowitz. Academic Press.

Hofstede, Geert H. 1980. *Culture's Consequences: International Differences in Work-Related Values.* Sage.

Jackson, Susan E. 1992. "Team Composition in Organizational Settings." Pp. 138-73 in *Group Process and Productivity,* edited by Stephen Worchel, Wendy Wood, and Jeffry Simpson. Sage.

Jacobs, Jerry A. 1992. "Women's Entry into Management: Trends in Earnings, Authority, and Values among Salaried Managers." *Administrative Science Quarterly* 37:282-301.

Kanter, Rosabeth M. 1977a. *Men and Women of the Corporation.* Basic Books.

———. 1977b. "Some Effects of Proportions on Group Life: Skewed Sex Ratios and Responses to Token Women." *American Journal of Sociology* 82:965-90.

Lehmann, Phyllis N. 1982, December. "Women in Orchestras: The Promise and the Problems." *Symphony Magazine.*

Lie, Suzanne S., and Virginia E. O'Leary. 1990. "In the Same Boat? Academic Women around the World." Pp. 17-29 in *Storming the Tower: Women in the Academic World,* edited by Suzanne Lie and Virginia E. O'Leary. Kogan Page.

Martin, Patricia Y. 1985. "Group Sex Composition in Work Organizations: A Structural-Normative Model." *Research in the Sociology of Organizations* 4:311-49.

Martin, Patricia Y., and Steve Harkreader. 1993. "Multiple Gender Contexts and Employee Rewards." *Work and Occupations* 20:296-336.

Moreland, Richard L., and John M. Levine. 1992. "The Composition of Small Groups." *Advances in Group Processes* 9:237-80.

Morgan, Ben B., Jr., and Donald L. Lassiter. 1992. "Team Composition and Staffing." Pp. 75-100 in *Teams: Their Training and Performance,* edited by Robert W. Swezey and Eduardo Salas. Ablex.

Nemeth, Charlan J., and Barry M. Staw. 1989. "The Tradeoffs of Social Control and Innovation in Groups and Organizations." Pp. 175-210 in *Advances in Experimental Social Psychology.* Vol. 22, edited by Leonard Berkowitz. Academic Press.

Pfeffer, Jeffrey. 1983. "Organizational Demography." Pp. 299-357 in *Research in Organizational Behavior.* Vol. 5, edited by Larry L. Cummings and Barry M. Staw. JAI Press.

Pfeffer, Jeffrey, and Alison Davis-Blake. 1987. "The Effect of the Proportion of Women on Salaries: The Case of College Administrators." *Administrative Science Quarterly* 32:1-24.

Previn, André (ed.). 1979. *The Orchestra.* Doubleday.

Reskin, Barbara F., and Patricia A. Roos. 1990. *Job Cues, Gender Cues: Explaining Women's Inroads into Male Occupations.* Temple University Press.

Ridgeway, Cecilia L., and David Diekema. 1992. "Are Gender Differences Status Differences?" Pp. 157-81 in *Gender, Interaction, and Inequality,* edited by Cecilia L. Ridgeway. Springer-Verlag.

Schmidt, Manfred G. 1992, September. *Gendered Labor Force Participation: Political and Social Determinants of Female and Male Rates of Labor Force Participation in 18 OECD Countries.* Paper presented at a conference on Comparative Studies of Welfare State Development, Bremen, Germany.

Smith, Kenwyn K. 1982. *Groups in Conflict.* Kendall/Hunt.

South, Scott J., Charles M. Bonjean, William T. Markham, and Judy Corder. 1982. "Social Structure and Intergroup Interaction: Men and Women of the Federal Bureaucracy." *American Sociological Review* 47:587-99.

Spangler, Eve, Marsha A. Gordon, and Ronald M. Pipkin. 1978. "Token Women: An Empirical Test of Kanter's Hypothesis." *American Journal of Sociology* 84:160-70.

Taylor, Shelley E., Susan T. Fiske, Nancy L. Etcoff, and Audrey J. Ruderman. 1978. "Categorical and Contextual Bases of Person Memory and Stereotyping." *Journal of Personality and Social Psychology* 36:778-93.

Tolbert, Pamela S., Tal Simons, and Alice Oberfield. 1994. *In Mixed Company: The Effects of Group Proportions on Group Dynamics.* Working Paper, School of Industrial and Labor Relations, Cornell University.

Toren, Nina. 1990. "Would More Women Make a Difference? Academic Women in Israel." Pp. 74-85 in *Storming the Tower: Women in the Academic World*, edited by Suzanne S. Lie and Virginia E. O'Leary. Kogan Page.

Touhey, John C. 1974. "Effects of Additional Women Professionals on Ratings of Occupational Prestige and Desirability." *Journal of Personality and Social Psychology* 29:86-89.

Tsui, Anne S., Terri D. Egan, and Charles A. O'Reilly. 1992. "Being Different: Relational Demography and Organizational Attachment." *Administrative Science Quarterly* 37:549-79.

Weiss, Howard M., and Seymour Adler. 1984. "Personality and Organizational Behavior." Pp. 1-50 in *Research in Organizational Behavior*. Vol. 6, edited by Barry M. Staw and Larry L. Cummings. JAI Press.

Wharton, Amy S., and Baron, James N. 1987. "So Happy Together? The Impact of Gender Segregation on Men at Work." *American Sociological Review* 52:574-87.

____. 1991. "Satisfaction? The Psychological Impact of Gender Segregation on Women at Work." *The Sociological Quarterly* 32:365-87.

Wood, Wendy, and Nancy Rhodes. 1992. "Sex Differences in Interaction Style in Task Groups." Pp. 97-121 in *Gender, Interaction, and Inequality*, edited by Cecilia L. Ridgeway. Springer-Verlag.

Yoder, Janice D. 1991. "Rethinking Tokenism: Looking beyond Numbers." *Gender and Society* 5:178-92.

Yoder, Janice D., Jerome Adams, and Howard Prince. 1983. "The Price of a Token." *Journal of Political and Military Sociology* 11:325-37.

Zanna, Mark P., Faye Crosby, and George Lowenstein. 1987. "Male Reference Groups and Discontent among Female Professionals." Pp. 28-41 in *Women's Career Development*, edited by Barbara A. Gutek and Laurie Larwood. Sage.

Social Structure, Political Institutions, and Mobilization Potential

RORY MCVEIGH, *University of North Carolina at Chapel Hill*

Abstract

The increasingly diverse set of issues that form the basis of both traditional and nontraditional political activity has been attributed to the structural changes occurring in postindustrial societies. In the social movement literature, surprisingly little attention has been paid to the specific ways in which social structure, rather than structural change, impacts interest formation and interest articulation. In this article I argue that some forms of structural differentiation promote distributions of interests in which some interests cannot effectively be channelled into electoral politics. Variables measuring the structural differentiation of Colorado counties, in income inequality, educational inequality, ethnic heterogeneity, and industrial heterogeneity, are used to examine the relationship or the "fit" between party voting and voting on ballot initiatives concerning social or cultural issues.

Social movement theorists in the U.S. have focused considerable attention on the question of how a collectivity, given a common interest, organizes in pursuit of a collective good. Meanwhile, their European counterparts have primarily focused on group grievances and on the way in which structural change generates new mobilization potentials, that is, on groups of individuals sharing common interests that could be potentially activated by social movement organizers (see Cohen 1985; Klandermans 1984).

Renewed interest in group grievances and interest formation can be attributed to the increasingly diverse set of issues that form the basis of both traditional and nontraditional political activity. Many contemporary movements, such as those promoting women's rights, the environment, gay rights, and pro-life and pro-choice concerns, are not strictly engaged in distributive battles, but instead are promoting values or a collective identity. The proliferation in recent decades of social movement organizations that are not engaged in distributive battles, and that therefore do not fit neatly onto a one-dimensional left-right

I am grateful to Peter Bearman, Judith Blau, Peter Blau, Ken Bollen, Kathleen Harris, Garry Hoefle, François Nielsen, Tony Oberschall, Rachel Rosenfeld, Mim Thomas, and anonymous reviewers for their useful comments and suggestions. ***Direct correspondence to Rory McVeigh, Department of Sociology, Hamilton Hall, CB #3210, The University of North Carolina, Chapel Hill, NC 27599-3210.***

 Social Forces, December 1995, 74(2):461-485

continuum, forces social movement theorists to once again consider the origins of group grievances and interests.

In this article I explore the role of social structure in the promotion of interest distributions, which refer to how political preferences on various issues are distributed within a given community. I argue that some forms of structural differentiation within a community promote interest distributions where nonmaterial interests cannot be effectively channeled into established political institutions. The argument is tested through an analysis of party voting and voting on ballot initiatives concerning social or cultural issues in Colorado from 1980 through 1990.

INTERESTS AND SOCIAL CLASS

Studies of voting behavior in the U.S. have consistently shown a correlation between social class and party voting. Those of lower economic status tend to vote for the Democratic party, while wealthier individuals are more likely to vote Republican (Lipset 1981). Factors other than class, such as religion, ethnicity, and education, have tempered intergroup antagonism. The absence of one deep societal cleavage based solely on class position has contributed to political stability and reinforced the democratic process (Blau & Schwartz 1984; Coleman 1957; Coser 1956; Lipset 1981; Rae & Taylor 1970).

Lipset, and other students of politics, have argued that in recent years the ties between class position and party voting have weakened (Lipset 1981; see also Brint 1985; Glenn 1973; Knoke 1976; Weakliem 1991; Weiner & Eckland 1979). According to Lipset, there are now two lefts, the "materialist" and the "postmaterialist." The materialist left advocates policies promoting economic equality while the postmaterialist left is primarily concerned with cultural issues, values, or collective identities.

Similar statements have been made by researchers studying "new social movements."[1] Offe (1985), for example, claims that until the mid-seventies, "the traditional left-right continuum was an approximately adequate model in which all relevant political and societal collective actors could be located" (857). He, along with others, has argued that in postindustrial societies, social and cultural issues have become more important than economic issues. Inglehart (1987) links societal affluence in advanced industrial countries with a greater freedom to be concerned with values, rather than material well-being. In a similar vein, Offe (1985) claims that previously unarticulated interests which were once subordinated to class struggle within the political sphere are now expressed by increasingly large "decommodified" segments of the population.

In short, new social movement theorists have argued that the distinction between state and society has become blurred. Political institutions have become less and less capable of incorporating the multitude of interests that result from structural transformations of advanced industrial societies. The predominant role of the new, decommodified middle class in new social movements is offered as supporting evidence of arguments stressing structural change.

Clearly, the prevalence of contemporary movements is related to structural changes in modern societies. Support for such movements, no doubt, is also related to what Lipset (1981) describes as the social bases of politics. But less

attention has been paid to the effects of social structure itself. Does the way in which a community is differentiated in income, education, ethnic, and occupational distributions have an impact on interest formation that is independent of the levels of income, wealth, or education? What types of social structure promote interest distributions that cannot be channeled into established political institutions?

Interests and Institutions

The first question that arises is, why are interests, whether they are instrumental or noninstrumental, pursued through a social movement rather than through traditional political institutions. In regard to class-based movements, it has been argued that negatively privileged groups are denied access to political power and therefore must rationally pursue their interests through nontraditional political means (Gamson 1975; Jenkins 1985; Jenkins & Perrow 1977; McAdam 1982; McCarthy & Zald 1977; Oberschall 1973; Piven & Cloward 1979; Tilly 1978). But what about movements centered around social or cultural issues? The membership for many of these movements, for example the environmental or pro-choice movements, is disproportionately drawn from the middle class (Cohen 1985; Offe 1985; Weakliem 1991). If the polity is responsive to groups possessing resources, why do these groups pursue their goals through social movement activity rather than through traditional politics?

Skocpol (1992) calls attention to the "fit" between a group's political orientations and the established political institutions designed to incorporate those interests; what we might call the "overall structure of political opportunity" (Gamson 1975; Jenkins 1985; McAdam 1982; Tilly 1978). When there is a poor fit between a group's interests and the capacity of established political institutions (such as political parties) to address those interests, social movement activity might be expected. Shefter (1983) argues, for example, that "the very process of channelling political participation through established institutions and into the electoral arena contains participation and limits its disruptive effects" (483; see also Amenta & Zylan 1991; Clemens 1993). "Disruption," therefore, should occur when interests cannot be channelled through the electoral arena.

INCONSISTENT INTERESTS

Electoral politics forces individuals to attempt to reconcile all the interests that they hold with the platform of one political party, a task that is often difficult to accomplish. At the national level, the Democratic party appeals to the left on both social and economic issues, while the Republican party tends toward the right in both spheres.[2] That is, the Democratic platform, relative to the Republican platform, advocates more state intervention in the economy to promote equality. On social issues, the Republican platform promotes traditional values and beliefs, while the Democratic platform appeals to postmaterialist values (Lipset 1981).

The ideological positions adopted by the two parties are consistent in the sense that the Democratic party positions itself to the left of center on both economic and social issues and the Republican party is right of center on both

types of issues. Interests held by individuals or groups, however, may or may not be consistent. While some in the working class are drawn toward conservative parties based on their beliefs in traditional cultural values, parties of the left pick up voters from the middle class who favor liberal policies on issues associated with lifestyle, culture, and values (Brint 1984; Lipset 1981; Weakliem 1991).

When a voter's interests are consistent, those interests can be reconciled with the position taken by one of the two competing parties. There is a fit between interests and political institutions, hence there is a low demand for political action that circumvents the traditional political arena. Voters with inconsistent interests, however, must rank their interests by preference, voting for the party that maximizes their utility (Schattsneider 1960). In doing so, some of their interests cannot be satisfactorily channelled into electoral politics. In fact voters are, in effect, forced to vote against their own position on some of the interests that they hold. While such cross-pressures may indeed reduce the antipathy that individuals feel toward the other party (Lipset 1981), they also generate a large pool of individuals who should be receptive to organizations pursuing their "unrepresented" interests outside of the traditional political arena.

For social movement theorists, therefore, it should be important to understand what factors contribute to interest inconsistency. When a desired collective good has jointness of supply, which is the case when the collective good in question concerns values or identity, effective collective action does not require full participation from potential beneficiaries (Oberschall 1993). A critical mass of highly motivated actors can do the job, and as Marwell and Oliver (1993) have demonstrated, the larger the size of the mobilization pool, the more likely it is that such a critical mass will emerge. For postmaterialist social movements, a large mobilization pool should exist in communities where liberal sentiment on social issues substantially exceeds support for the Democratic party. Similarly, a large mobilization pool for conservative, nondistributive movements should exist where conservatism on social issues exceeds support for the Republican party.

Social Structure: Inequality and Heterogeneity

Drawing on the work of Peter Blau, I propose that interest inconsistency is related to variations in the social structure of local environments. Blau has written extensively on the impact of social structure on social relations. He argues that individual choices and preferences are constrained, to a great extent, by the social structure in which individuals are embedded. Social structure, using Blau's definition, is "the distribution of a population among positions in a multidimensional space" (Blau & Schwartz 1984:9).

Societies are differentiated by nominal distinctions such as ethnicity or occupation, but they are also differentiated by status gradations such as income or levels of education. At the individual level, a considerable amount of research has shown that people tend to associate with others who are most similar to themselves, in terms of both nominal and graduated parameters

(Abramson 1973; Alba 1976; Blau & Duncan 1967; Heer 1974; Hollingshead 1950). Paradoxically, Blau shows that at the societal level heterogeneity promotes intergroup relations and inequality increases the probability of status-distant relations. In a completely homogeneous society, for example, there is a zero probability of having intergroup relations, and when there are high levels of equality, there will be few status-distant relations.

These structural constraints should have an important impact on how views on economic policy and social policy coincide within a given community. In a community with little inequality, there is little contact between individuals of different status. Holding constant the absolute level of resources, the promotion of equality is unlikely to be a salient issue when inequality is low. Where there is high inequality, however, there are higher levels of status-distant contact. A minority of the population possesses a large share of the wealth or income, and the majority is relatively impoverished. The presence of inequality is likely to contribute to feelings of deprivation or injustice held by the underprivileged majority, generating a demand for redistributive policy (see Jasso 1980).

Differentiation on nominal parameters should have implications for the demand for public policies reflecting postmaterialist values. Liberalism on social issues addresses the rights of individuals, for example, the right of a woman to choose to have an abortion, or the right of a gay individual to serve in the military. Social conservatism, however, encourages conforming to established tradition or to the dominant values of society. Heterogeneity on nominal parameters within an individual's immediate environment, I argue, promotes individuality, while homogeneity promotes conformity.

As Blau and Schwartz state, cross-cutting cleavages that are characteristic of heterogeneous communities "produce a complex web of group affiliations, and individuals find themselves at the intersection of numerous groups. This creates cross pressures, which are stressful, but which also weaken the power of a given group to enforce restrictions on individuals, thereby expanding freedom of choice" (Blau & Schwartz 1984:84). In a homogeneous community, on the other hand, a limited role structure exerts pressure on the individual to conform to the dominant values of his or her immediate environment (Blau & Schwartz 1984; Durkheim [1933] 1984; Simmel 1950; Thomas 1994).

If the preceding logic is sound, it would seem that both inequality and heterogeneity should promote Democratic voting. In reality, however, we see many heterogeneous and unequal communities that vote Republican.[3] This is consistent with the arguments made by Lipset and by new social movement theorists that in modern industrial societies, cultural issues are becoming at least as important as class issues in the political arena. For this reason, it seems particularly important that we address the relationship between nominal and graduated parameters in the study of political interests.

Uncovering the net impact of inequality, however, requires holding constant the levels of heterogeneity. In the same manner, to assess the net impact of heterogeneity, levels of inequality must be held constant. The degree that nominal and graduated parameters intersect varies considerably from community to community. According to the argument made above, where there is both inequality and heterogeneity, there should be consistency. This type of structure fosters liberal sentiment in both the economic and social dimensions that can be

expressed through the Democratic party. There should also be consistency where there is homogeneity and equality. In this case the structural parameters do not foster a demand for either liberal economic policy or liberal social policy. This interest distribution can be effectively captured by the Republican party.

It is where the structural parameters work at cross-purposes that interest inconsistency arises. For instance, where there is inequality and homogeneity there should be a demand for policies promoting economic equality, but little demand for social liberalism. Or, when a community is characterized by equality and heterogeneity, there should be support for economic conservatism, but a demand for liberal policies on social issues. Neither of these combinations can be reconciled with the platform of one of the two major parties. These predictions can be summarized in the two-way Table 1. I refer to communities where there is social conservatism but economic liberalism as "social reactive" and to those communities where there is social liberalism and economic conservatism as "social proactive."[4]

The Data

I use ballot initiative voting not as a measure of social movement activity, but instead as a way of measuring the fit between party voting and positions on social issues. Initiatives and referenda have increasingly been used as a means of pursuing and expressing interests outside of party politics (Hahn & Kamieniecki 1987). The costs and effort required to put an initiative on the ballot are small relative to the costs entailed in starting a new and viable political party. Initiatives allow individuals or groups to bypass traditional politics, putting their position before the general public. They provide voters holding inconsistent interests with the freedom to remain within the party based on one set of preferences, but act outside of the party on other interests.

In the 1980s, several initiatives appeared on Colorado ballots centering on issues that cannot be considered as strictly distributive battles. All these issues have been, in some form, addressed in the social movement literature. They include abortion (Clarke 1987; Eckberg 1988; Leahy & Mazur 1978; Luker 1984; Staggenborg 1988, 1989), gambling or vice (Beisel 1990; Gusfield 1963; Wood & Hughes 1984), the environment (Barbosa 1993; Mitchell 1979; Modavi 1991; Offe 1990), restrictions on the sale of alcohol (Gusfield 1963; Unnithan & Corzine 1987), and multilingualism (Laczko 1987; Nielsen 1980; Olzak 1982). These issues, of course, also contain an economic component, but in each one, culture, identity, and ideology play central roles.

Counties in Colorado are used as units of analysis; they reflect considerable structural diversity in income, educational levels, occupational distributions, and ethnic composition. I include eight ballot initiatives in the analysis, and these represent all the initiatives appearing on the November ballot in Colorado from 1980 through 1990 that concern social or cultural issues. Democratic voting is operationalized as each county's percentage of votes for the Democratic candidate for the U.S. Senate, averaged across the four elections held from 1980 to 1990.[5] A description of the eight initiatives and their outcomes is provided in

TABLE 1: Structural Differentiation and the Demand for Liberal or Conservative Social and Economic Policy

	Homogeneity	Heterogeneity
Inequality	Social Conservative and Economic Liberal (Social Reactive)	Social Liberal and Econimic Liberal (Consistent)
Equality	Social Conservative and Economic Conservative (Consistent)	Social Liberal and Economic Conservative (Social Proactive)

Table 2. Pearson correlations and univariate statistics for Democratic voting and the initiative results are shown in Table 3.

The correlations reported in Table 3 show that the liberal position on ballot initiatives tends to be associated with Democratic voting. This is the case in seven of the eight initiatives used in this study. In the eighth initiative (gambling in 1990), party voting has no correlation with proposition voting. There is a particularly strong positive correlation between Democratic voting and support for bilingualism ($r = .688$).

In addition to party voting and initiative voting, there are several relatively strong correlations among the voting results of the initiatives. As would be expected, there is an extremely high correlation between the two abortion initiatives (.921). There is also a high positive correlation between the proabortion stance and the prolottery stance (.738). Voting "yes" on the proalcohol initiative is highly correlated with voting "yes" on the proenvironment issue (.707). Other correlations among the ballot initiatives are weaker.

While liberal voting on ballot initiatives does tend to be associated with Democratic party voting, the fit is far from perfect. This study uses the voting data that have been presented to form a measure of interest inconsistency. Variables reflecting structural differentiation of counties are then used to examine the relationship between structural differentiation and interest inconsistency.

THE DEPENDENT VARIABLE

A ratio measure is used to capture the fit, or lack of fit, between interests on social issues and party voting. For each ballot initiative, the county's percentage vote for the liberal stance is first divided by the sample mean. I then divide this ratio by the county's Democratic voting percentage divided by the sample mean. This ratio of two ratios provides a measure of the extent to which a county's support for the liberal position on ballot initiatives exceeds or is less than its support for the Democratic party. The numerator measures the extent to which a county favors the liberal position on a given social issue, relative to

TABLE 2: All Colorado Initiatives Concerning Social Issues That Appeared on November Ballots from 1980-1990

Initiative	Sponsor's Position	Outcome	Description
Amendment 2, 1980	Liberal	59.9% in favor	Amendment to change state constitution's antigambling clause, to allow for legislative creation of a state lottery
Amendment 5, 1982	Liberal	26.5% in favor	Would require 5 cent deposit on beverage bottles and cans. It would also ban detachable pop tops and nonbiodegradable ring carriers.
Amendment 7, 1982	Liberal	34.6% in favor	Would allow grocery stores to sell wine.
Amendment 3, 1984	Conservative	50.4% in favor	Would prohibit publically funded abortions.
Amendment 5, 1984	Liberal	33.6% in favor	Would permit casino gambling in Pueblo, Colorado.
Amendment 1, 1988	Conservative	60.8% in favor	Would make English the official language in Colorado, and would place restrictions on speaking other languages, for government employees.
Amendment 7, 1988	Liberal	39.8% in favor	Would repeal ban on using state money for abortions.
Amendment 4, 1990	Liberal	57.6% in favor	Would allow limited stakes gambling in two historic mining towns in Gilpen County.

[a] Source: *Denver Post*

TABLE 3: Pearson Correlations and Univariate Statistics for Democratic Voting and the Liberal Position on Initiative Voting

	Democrat	1	2	3	4	5	6	7	8
Democratic voting	1								
(1) Prop. 2 1980 (prolottery)	.321	1							
(2) Prop. 5 1982 (proenvironment)	.38	.493	1						
(3) Prop. 7 1982 (proalcohol)	.184	.635	.707	1					
(4) Prop. 3 1984 (proabortion funding)	.361	.7	.594	.547	1				
(5) Prop. 5 1984 (progambling)	.281	.193	.064	.186	-.075	1			
(6) Prop. 1 1988 (against English only)	.688	.006	.295	.06	-.022	.351	1		
(7) Prop. 7 1988 (proabortion funding)	.334	.738	.659	.67	.921	-.033	-.034	1	
(8) Prop. 4 1990 (progambling)	.001	.258	-.224	-.02	-.008	.416	.009	-.025	1
Mean	40.27	53.47	24.2	32.25	45.54	35.78	37.92	34.52	56.68
Std. dev.	7.9	11.64	8.81	11.4	8.9	7.52	12.16	10	5.94
Minimum	28.97	31.5	10.3	15.5	31	24	17	16	44.1
Maximum	63.51	79.6	50	64.9	73.1	67.2	83.3	62.2	70.8

(N=63)

[a] Colorado counties are the units of analysis.

the vote of other counties. The denominator is a measure of the relative support for the Democratic party.[6] A high value on the overall measure, then, signifies that the county is more liberal on the social issue in question than would be expected based on its party voting (social proactive). A low value, on the other hand, means that the county is more conservative on social issues than party voting would suggest (social reactive). Values close to 1 indicate consistency between party voting and voting on initiatives.

INDEPENDENT VARIABLES

Four variables are used as measures of the structural differentiation of the counties: income inequality, educational inequality, ethnic heterogeneity, and industrial heterogeneity. I also control for median income, median years of education, party competitiveness, and population density. See Table 4 for the method of calculation for each variable and the source of the data.

Income Inequality and Educational Inequality

Based on the previous discussion, I expect that both income inequality and educational inequality will be negatively related to social proactive voting. I have argued that inequality on graduated parameters promotes a demand for liberal economic policy, but not necessarily liberal social policy. Holding the counties' level of heterogeneity constant, therefore, inequality should be associated with more social conservatism than would be expected based on party voting, while equality should be associated with more liberalism on social issues than party voting would suggest.

Industrial Heterogeneity and Ethnic Heterogeneity

As previously argued, heterogeneity on nominal parameters such as ethnicity and occupation should have a positive impact on social proactive voting. The presence of many cross-cutting cleavages fosters individualism, which is consistent with liberal or postmaterialist values, while a limited role structure exerts pressure towards conformity. Holding constant the level of inequality, heterogeneity should be associated with higher levels of social liberalism relative to Democratic voting, and homogeneity should be associated with lower support for social liberalism relative to Democratic voting.

CONTROL VARIABLES

Party Competitiveness

I control for party competitiveness since environments in which one party dominates should be conducive to the formation of new political parties purely from an ecological standpoint. In such a case, the dominant party occupies an expanded niche and is vulnerable to losing its peripheral supporters (see Nielsen 1986; Pinard 1975). Since there are several barriers to third party formation in the U.S., those interests that a third party might address, such as different combinations of liberalism and conservatism, might be expressed through other means. Since most of the counties in Colorado have Republican

TABLE 4: Definitions of Independent Variables and Sources of Data

Income inequality: Data on family income are taken from Table 180 of the 1980 U.S. Census of Population. The census presents the income distribution for Colorado counties in 9 categories. The categories are (1) less than \$5,000, (2) \$5,000-\$7,499, (3) \$7,500-\$9,999, (4) \$10,000-\$14,999, (5) \$15,000-\$19,999, (6) \$20,000-\$24,999, (7) \$25,000-\$34,999, (8) \$35,000-\$49,999, and (9) more than \$50,000. To calculate the Gini coefficient, a Pareto interpolation is first used to estimate the average income in categories above the category containing the median income (see Nielsen & Alderson 1993). The coefficient is then calculated using the formula provided by Nygard and Sandstrom (1981).

Educational inequality: In Table 175 of the 1980 census, years of education for persons 25 years and older is aggregated into 8 categories. The categories are (1) 0-4 years, (2) 5-7 years, (3) 8 years, (4) 9-11 years, (5) 12 years, (6) 13-15 years, and (8) 17 years or more. In calculating a Gini coefficient, the average is estimated as the midpoint of each category. I assume that the average of the open ended category is 19 years of education.

Industrial heterogeneity: Industrial heterogeneity is measured using the 11 primary industry categories of employment reported in the 1980 census, Table 178. These categories are (1) agriculture, (2) forestry and fishery, (3) mining, (4) construction, (5) manufacturing, (6) transportation, communications, and other public utilities, (7) wholesale trade, (8) retail trade, (9) finance, insurance, and real estate, (10) services, and (11) public administration. The heterogeneity measure is calculated as $1 - \Sigma P_i^2$, where P_i is the fraction of the population in each group (Blau 1977).

Ethnic heterogeneity: Data for ethnic heterogeneity comes from table 178 of the 1980 census. Five categorical groupings are used. They are (1) white (not Spanish origin), (2) black (not Spanish origin), (3) Spanish origin, (4) American Indian, and (5) Asian. The heterogeneity measure is calculated in the same manner as for industrial heterogeneity.

Party competitiveness: I use the earlier described measure of voting for U.S. Senate campaigns to construct a measure of party competitiveness. The measure is constructed in the same manner as the heterogeneity variables. Only two categories, Democratic voting and Republican voting, are used. In Colorado, support for any other party in senate campaigns was either minimal, or nonexistent.

Population density: Population density is measured with the natural log of the population per square mile, as reported in Table 2 of the 1980 census.

Median income: County-level median income is taken from Table 189 of the U.S. census.

Median years of education: County-level median years of education is taken from Table 175 of the U.S. census.

majorities, party competitiveness is likely to be negatively related to proactive voting. That is, in most cases noncompetitiveness means a large Republican majority, thereby placing a limit on the extent that conservative initiative voting can exceed voting for the Republican party.

Population Density

Population density captures the extent that counties are urbanized. Population per square mile is also related to opportunity for contact — a key assumption of Blau's (1977) theory. Any type of social relations, favorable or hostile, are dependent upon propinquity.

Median Income

To ascertain the impact of structural differentiation on interest inconsistency, it is necessary to control for absolute levels of wealth or income of the counties. It could be argued, for example, that measures of income inequality merely serve as a proxy for the relative wealth or poverty of the counties. Several voting studies have called attention to the phenomenon of upper-status economic conservatism and social liberalism (Brint 1984; Hamilton 1975; Huber & Form 1973; Lipset 1981; Robinson & Bell 1978). It should be expected that median income will have a positive impact on proactive voting.

Median Years of Education

It is particularly important to control for the median level of years of education since a large body of research has documented a strong positive relationship between education and liberalism on social issues (see Hyman & Wright 1979; Lipset 1981; Nunn et al. 1978).[7]

Methods

I begin by examining the relationship between social structure and the broad concept of interest inconsistency. Subsequently I focus on the extent to which these structural variables contribute to our understanding of interest inconsistency on the specific issues. More emphasis is placed on the broader concept. Political representatives, after all, must attempt to address a wide array of constituent interests and preferences. Analyzing all eight issues simultaneously will shed light on how structural differentiation impacts the fit between party voting and the many dimensions of social liberalism. Nevertheless, it is also useful to examine the relationship between social structure and interest inconsistency on the specific issues. For example, do measures of inequality and heterogeneity help us understand why some counties have strong support for the Democratic party, yet strongly oppose abortion funding?

A panel design is used to analyze all the initiatives simultaneously. Essentially, the ballot initiative votes represent eight cross-sections for the 63 Colorado counties. The dependent variable in each cross-section is the measure of the counties' social proactiveness. To control for possible effects of the substance of a ballot initiative on proactiveness, I use a fixed effects model. In this approach, initiative specific dummy variables corresponding to each panel

are inserted, measuring the shifts in the regression line that may result from unknown variables associated with the specific issues (see Kennedy 1992:222). The model is estimated as

$$Y_{it} = \alpha_1 d_{2it} + \ldots + \beta' X_{it} + \epsilon_{it}$$
$$= \alpha + \beta' X_{it} + \epsilon_{it}$$

where Y_{it} is the proactiveness of county i on ballot t, X_{it} is a vector of independent variables for county i at ballot t, β is the vector of regression coefficients, the α_i's are initiative specific constants, and the d_j's are the eight issue specific dummy variables which equal 1 only when $j = i$. (Greene 1992).[8]

The matrix of Pearson correlation coefficients in Appendix A shows that there are high correlations between the income and education control variables and the educational inequality measure. The zero order correlation between median income and educational inequality is -.795, and the correlation between median years of education and educational inequality is -.819. Therefore, three models are presented. In the first model, only the control variables are used. In the second model, the structural variables are added but median income and median education are omitted. In the third model, all the variables are included.

After examining the impact of structural differentiation on proactive voting when all of the issues are considered simultaneously, I apply the full model to each initiative. Ordinary least squares regression is used to estimate eight separate models. The dependent variable for each model is calculated in the same manner as in the panel design, capturing the extent that liberal voting on the social issue exceeds support for the Democratic party.

Estimating each vote separately substantially reduces the degrees of freedom. This could be a problem since median income, median education, and educational inequality are highly correlated. As a conservative test on the structural variables, however, I include both median education and median income in the model. The effects of the structural variables, after all, are likely to be associated with the absolute levels of income and education.

The Findings

THE BROAD CONCEPT OF SOCIAL LIBERALISM

The results of the pooled models are reported in Table 5. In model 1, 21% of the variation in social proactive voting is explained by the control variables. Party competitiveness has the expected negative impact on social proactiveness. In other words, the more party competitiveness within the county, the less support for social liberalism relative to support for the Democratic party. Both median income and median years of education have the anticipated positive effect.[9]

In the second model, where the structural variables are added and median income and education are left out, the R^2 increases to .280. Income inequality, which I expected to have a negative effect on proactiveness, is not statistically significant, although the sign of the coefficient is in the expected direction. As was the case in the first model, party competitiveness is negatively related to social proactiveness. Population density also has a strong negative effect,

perhaps because the more urbanized areas tend to have the strongest support for the Democratic party. After controlling for ethnic and industrial heterogeneity, which are also typical of urbanized areas, there is less support for social liberalism relative to support for the Democratic party in densely populated counties.

As predicted in Table 1, ethnic and industrial heterogeneity are both positively related to social proactive voting. Holding the other variables constant, the more heterogeneous the county, the greater the support for social liberalism relative to Democratic voting. The more homogeneous the county, the more support for social conservatism, relative to support for the Republican party. Educational inequality also has the anticipated negative impact on social proactive voting. Net of the effects of the other variables, educational inequality promotes less support for social liberalism relative to support for the Democratic party. Educational equality, on the other hand, fosters more social liberalism relative to Democratic voting.

In the full model, where all the control variables are included, median income and median education are no longer significant at the .05 level (most likely due to the high correlation between these variables). The inclusion of these controls, however, does not eliminate the statistical relation between the structural variables and social proactive voting. Once again, income inequality is not statistically significant. The relationships for the other variables remain strong. Party competitiveness and population density are negatively related to social proactive voting, as is educational inequality. Occupational heterogeneity and ethnic heterogeneity have the expected positive relation.[10]

The partial coefficient for the ethnic heterogeneity variable calls attention to the importance of the control variables in each model. The zero-order correlation between proactive voting and ethnic heterogeneity is negative ($r = -.236$), yet the results of multivariate analysis show the predicted positive relationship between these variables. Since ethnic heterogeneity is at least moderately correlated with most of the variables in the analysis, multivariate analysis is needed to assess the net impact of this variable. For example, ethnically heterogeneous counties tend to have lower median income and median education. They also tend to have more party competitiveness and educational inequality.

While it is important to control for these other variables, the strong correlations between some of the variables introduces a potential problem of multicollinearity. To assess the extent that the findings are robust, I begin by running a model without the control variables. As shown in Table 6, educational inequality and industrial heterogeneity are highly significant in the predicted direction. Income inequality is not significant, and the coefficient for ethnic heterogeneity, while no longer negative, is insignificant.

Next, I present several models including three of the four control variables (I drop median income from the analysis due to its high correlation with median education). Then, controlling for population density, party competitiveness, and median years of education, I add the structural variables to the model one at a time, beginning with ethnic heterogeneity.

As shown in the second column of Table 6, the relationship between ethnic heterogeneity and social proactive voting becomes positive, and has borderline

TABLE 5: Fixed Effects Model Estimates: Proactive Voting in Colorado Counties on Eight Ballot Initiatives

	1	2	3
Control variables			
Party competitiveness	-4.127*** (.5013)	-4.6533*** (.5530)	-5.1429*** (.6127)
Population density (logged)	-.0062 (.0080)	-.0287*** (.0080)	-.0279*** (.0090)
Median income (thousands)	.01073*** (.0038)	—	.0021 (.0049)
Median education	.0928*** (.0216)	—	.0559* (.0314)
Structural variables			
Income inequality	—	-.207 (.2679)	-.399 (.299)
Educational inequality	—	-4.925*** (.4575)	-3.862*** (.7771)
Industrial heterogeneity	—	1.6259*** (.3768)	1.5974*** (.3846)
Ethnic heterogeneity	—	.7772*** (.1238)	.7574*** (.1243)
R^2	.212	.28	.285
Adusted R^2	.194	.261	.263

[a] Standard errors are in parentheses. Issue-specific dummy variables are included.

	1	2	3
Joint significance of issue specific dummy variables	F = .176 df = 7/492 p = .989	F = .185 df = 7/490 p = .988	F = .186 df = 7/488 p = .988
Joint significance of structural variables	— — —	F = 32.33 df = 4/490 p < .01	F = 12.456 df = 4/488 p < .01

* p < .10 ** p < .05 *** p < .01

TABLE 6: Fixed Effects Model Estimates: Proactive Voting in Colorado Counties on Eight Ballot Initiatives

	1	2	3	4	5
Control variables					
Party competitiveness	—	-5.0460***	-5.7442***	-5.605***	-5.1260***
	—	(.5890)	(.6295)	(.6262)	(.6108)
Population density (logged)	—	.00085	.00109	-.0046	-.0263***
	—	(.0073)	(.00724)	(.0074)	(.0081)
Median education	—	.15619***	.1769***	.1941***	.0566*
	—	(.0202)	(.0212)	(.0218)	(.0314)
Structural variables					
Income inequality	.3259	—	-.8767***	-.8977***	-.4238
	(.2810)	—	(.0212)	(.2917)	(.2931)
Educational inequality	-2.6740***	—	—	—	-4.0200***
	(.3935)	—	—	—	(.2931)
Industrial heterogeneity	1.4560***	—	—	1.1328***	1.6310***
	(.3850)	—	—	(.3790)	(.3759)
Ethnic heterogeneity	.02666	.1892*	.2633***	.3140***	.75524***
	(.09821)	(.0998)	(.1021)	(.0218)	(.1241)
R^2	.147	.205	.219	.233	.285
Adusted R^2	.128	.187	.200	.213	.274

[a] Standard errors are in parentheses. Issue-specific dummy variables are included.

Joint significance of issue specific dummy variables.

F = .162	F = .173	F = .175	F = .178	F = .186
df = 7/492	df = 7/492	df = 7/491	df = 7/490	df = 7/489
p = .992	p = .991	p = .990	p = .99	p = .988

Joint significance of structural variables

—	F = 3.776	F = 6.377	F = 7.298	F = 1 4.704
—	df = 1/492	df = 2/491	df = 3/490	df = 4/489
—	p < .10	p < .01	p < .01	p < .01

* p < .10 ** p < .05 *** p < .01

significance ($p = .058$), after controlling for population density, party competitiveness, and median education. Interestingly, when income inequality is added to the model, as shown in model 3, not only does the impact of ethnic heterogeneity become stronger, but income inequality also has the anticipated negative impact on proactive voting ($p = .002$). When occupational heterogeneity is added in model 4, the observed effects of ethnic heterogeneity and income inequality remain intact, and occupational heterogeneity has the predicted positive effect. It is when educational inequality is added to the model (model 5) that the relationship between income inequality and proactive voting is no longer statistically significant.

THE SPECIFIC ISSUES

In this section I examine the extent to which the structural variables can explain proactive voting on each specific issue. As a conservative test, I include both median income and median education as controls. The results are reported in Table 7.

These findings show that with only three exceptions, out of a possible 32, the signs of the coefficients for the structural variables are in the expected direction. Contrary to my argument, income inequality has a positive sign on the environmental issue and on the bilingualism issue. Industrial heterogeneity is unexpectedly negative on gambling in 1990. The coefficients are in the expected direction for the other 29 possibilities. As for the control variables, the coefficient for party competitiveness is consistently negative. The coefficient for population density is also negative on all eight issues. Median education yields 6 positive coefficients, while median income yields 5.

The effects of the structural variables are particularly strong for the environment and alcohol votes. For these issues, income inequality is the only structural variable that is not statistically significant at the .05 level. At least one of the structural variables is significant at the .05 level in all the models, with the exception of abortion in 1984. Both educational inequality and industrial heterogeneity are significant predictors of proactive voting for the 1988 abortion initiative, but are not significant for the 1984 abortion vote. For abortion, the median levels of education and income are significant in 1984, while in 1988 they are not. A larger study population would be useful in separating out the effects of variables measuring absolute levels of income or education, and variables measuring the distribution of income and education. As in the pooled model, I examine the stability of the partial coefficients by controlling for party competitiveness, population density, and median education, and then adding each structural variable one at a time. The partial coefficients and probability values are reported in Appendix C.[11]

Holding all else constant, income inequality does not seem to have much of an effect on most of the issues. Income inequality is, however, a significant predictor for gambling in 1990. The educational inequality measure, on the other hand, is statistically significant at the .05 level on three of the eight issues, and is significant at the .10 level on gambling in 1988. Even after controlling for median education and median income, educational inequality has a negative impact on social proactive voting on the environmental issue, the alcohol issue,

TABLE 7: Ordinary Least Squares Estimates: Proactive Voting in Colorado Counties on Eight Ballot Initiatives

	Lottery 1980	Environ. 1982	Alcohol 1982	Abortion 1984	Gamblg. 1988	Bilingual. 1988	Abortion 1988	Gamblg. 1990
Constant	2.441**	1.404	3.022	1.986**	5.059***	0.739	0.874	0.6049***
	(1.027)	(1.696)	(1.986)	(.793)	(1.269)	(1.208)	(1.06)	(.797)
Control variables								
Party competitive.	-5.653***	-5.421***	-10.953***	5.6***	-6.737***	-1.958	-4.184***	-8.069
	(1.106)	(1.827)	(2.141)	(.854)	(1.367)	(1.302)	(1.142)	(.859)
Population density (logged)	-0.007	-0.037	-0.019	-0.027**	-0.066***	-0.042	-0.017	-0.001
	(.016)	(.027)	(.031)	(.012)	(.020)	(.019)	(.017)	(.012)
Median inc. (1,000s)	0.020**	-0.025*	-0.030*	0.023**	0.017	-0.006	0.008	0.007
	(.009)	(.015)	(.017)	(.007)	(.011)	(.067)	(.009)	(.007)
Median education	0.025	0.151	0.215*	0.098**	-0.115	0.018	0.098*	-0.048
	(.057)	(.094)	(.110)	(.044)	(.070)	(.067)	(.059)	(.044)
Structural variables								
Income inequality	-0.394	0.085	-1.730	-0.652	-0.340	0.878	-0.400	-1.325***
	(.538)	(.889)	(1.041)	(.416)	(.665)	(.633)	(.556)	(.418)
Educational inequality	-2.319	-9.171***	-9.542***	-0.591	-3.441*	-1.93	-4.174***	-0.854
	(1.411)	(2.331)	(2.730)	(1.090)	(1.744)	(1.661)	(1.457)	(1.096)
Industrial heterogeneity	1.417**	2.681**	3.88***	0.479	1.302	1.101	2.008	-0.111
	(.694)	(1.146)	(1.343)	(.536)	(.858)	(.817)	(.717)	(.539)
Ethnic heterogeneity	0.096	1.540***	1.550***	0.162	0.877***	1.308***	0.303	0.158
	(.226)	(.373)	(.437)	(.174)	(.279)	(.266)	(.233)	(.175)
R^2	0.684	0.532	0.578	0.730	0.532	0.450	0.708	0.751
Adjusted R^2	0.637	0.463	0.516	0.689	0.462	0.369	0.664	0.714
(n = 63)								

[a] Standard errors are in parentheses.

Joint significance of structural variables

	Lottery 1980	Environ. 1982	Alcohol 1982	Abortion 1984	Gamblg. 1988	Bilingual. 1988	Abortion 1988	Gamblg. 1990
F	F = 1.708	F = 5.625	F = 5.726	F = .95	F = 2.711	F = 7.56	F = 3.467	F = 3.036
df	df= 4/54	df = 4/54	df= 4/54	df = 4/54	df= 4/54	df= 4/54	df = 4/54	df = 4/54
p	N.S.	$p < .01$	$p < .01$	N.S.	$p < .05$	$p < .01$	$p < .05$	$p < .05$

* $p < .10$ ** $p < .05$ *** $p < .01$

the 1988 abortion issue, and the 1988 gambling issue. Industrial heterogeneity is significant at the .05 level on four of the eight issues. This variable has the expected positive impact on the lottery issue, the environment issue, alcohol, and the 1988 abortion issue. Finally, ethnic heterogeneity is positive and highly significant on the environment, alcohol, gambling, and bilingualism votes.

Discussion and Conclusion

The preceding analysis sheds light on the relationship between structural differentiation and interest distributions in the state of Colorado. The findings also point to several important research questions that could be pursued in future work. First of all, the results of this analysis suggest that it is important to consider the impact of structural differentiation on the formation and articulation of political interests. I have argued that inequality within a community fosters a demand for redistributive economic policy, while equality does not. Heterogeneity promotes a demand for liberal social policy, while homogeneity does not. When heterogeneity coincides with inequality within a community, or when homogeneity coincides with equality, interests can be effectively represented by the two competing political parties. When heterogeneity and inequality work at cross purposes, however, there should be a large pool of individuals who cannot channel all their interests into electoral politics.

In support of this argument the findings of this analysis show that holding the other variables constant, industrial heterogeneity and ethnic heterogeneity promote greater support for social liberalism relative to support for the Democratic party, when the eight initiatives are considered simultaneously. Also, the greater the educational inequality, the less support for social liberalism relative to support for the Democratic party. Income inequality, however, does not have a significant impact after controlling for educational inequality.

Looking at each initiative separately, it is clear that depending upon the issue, different structural variables come into play. The argument presented in this article receives the strongest support when considering the environmental issue and the alcohol issue, where educational inequality, occupational heterogeneity, and ethnic heterogeneity have the predicted effect. But on six of the eight initiatives, a test of the joint significance of the four structural variables is significant at least at the .05 level. The exceptions are the 1980 lottery vote and the 1984 abortion vote. Also, for seven of the eight initiatives, at least one of the structural variables is statistically significant and in the expected direction, the exception being abortion in 1984.

The failure of the structural variables to predict interest inconsistency on the 1984 abortion vote brings up an important point. On this issue, more than any of the others, religion plays a role that is not captured in the model. On the one hand, industrial heterogeneity intersects with gender, promoting a proactive gender identity at odds with a more traditional view of a woman's role in society (see Luker 1984). But in Colorado, ethnic heterogeneity usually means a high proportion of Mexican Americans, who are more likely than the other ethnic groups to have a religious affiliation (Catholicism) that is in conflict with the proabortion position. It is not surprising, in this case, that ethnic heteroge-

neity does not have the expected effect. Even so, it should be noted that on the 1988 abortion vote, educational inequality and industrial heterogeneity do play a role in explaining interest inconsistency.

Finally, the limited effectiveness of the income inequality variable is particularly interesting. While inequality does seem to affect interest inconsistency, it is educational inequality rather than income inequality that is important. Consistent with Lipset's description of a "postmaterialist left," this finding could reflect the declining significance of economic class in advanced industrial democracies and the increasing importance of cultural capital embodied in higher education. In the pooled model and in the models predicting interest inconsistency on the environment, alcohol, and the 1988 abortion issues, educational inequality had a highly significant negative impact on proactive voting. In other words, in those counties where a relatively small proportion of the population had access to higher education, there was more social conservatism than would be expected based on the proportion of the counties' vote for the Democratic party.

In such counties where interests are "inconsistent," there should be a mobilization pool that could be tapped by social movement activists pursuing nonmaterial interests. In some Colorado counties, popular sentiment is conservative on social issues, yet the county votes Democratic. In other counties, there is predominantly liberal sentiment on social issues, but the county nevertheless votes Republican. These counties would probably be receptive to movement organizations that provide a means for expressing unrepresented interests outside of traditional political institutions.

Social movements are one possible response to interest inconsistency. Other responses might be increased use of ballot initiatives and referenda, the formation of new political parties, or realignments within the dominant parties. Whatever the response, our understanding of both traditional and nontraditional political action should be enhanced by studying the relationship between social structure, political institutions, and mobilization potential.

Notes

1. While much of this research has originated in Europe, social movement organizations promoting values or collective identities are not missing from the American political landscape.

2. While political organization takes place at many levels, it is difficult for individuals to escape the impact of state or national level policies on issues such as abortion, the environment, gay rights, and economic policy.

3. In fact in this study of Colorado county level voting, I find no zero-order correlation between party voting and variables measuring inequality. Using the measures described later in the article, the correlation between Democratic voting and income inequality is -.011 and the correlation between Democratic voting and educational inequality is .054. Ethnic heterogeneity does show a strong positive correlation with Democratic voting (.447), but the correlation between Democratic voting and industrial heterogeneity is -.167.

4. Tilly refers to movements seeking to reassert established claims as reactive, while movements attempting to secure new benefits are proactive (Tilly 1978).

5. I use this measure, as opposed to party registration figures, because it provides a measure of how people actually vote when confronted with a choice between two candidates of the major competing parties.

APPENDIX A: Pearson Correlations and Univariate Statistics for All Variables Used in Pooled Model: 63 Colorado Counties at 8 Elections

	1	2	3	4	5	6	7	8	9
(1) Proactive voting	1								
(2) Median income (1,000s)	0.245	1							
(3) Median years education	0.245	0.659	1						
(4) Party competitive.	-0.29	0.176	0.269	1					
(5) Log of population density	-0.039	0.464	0.189	0.337	1				
(6) Income inequality	0.054	-0.093	0.165	-0.224	0.018	1			
(7) Educational inequality	-0.342	-0.795	-0.819	-0.037	-0.192	-0.016	1		
(8) Ethnic heterogeneity	-0.236	-0.392	-0.387	0.42	0.276	-0.036	0.632	1	
(9) Industrial heterogeneity	0.095	0.005	-0.233	-0.197	0.112	0.001	0.18	-0.054	1
Mean	1.016	18.1	12.82	0.469	2.151	0.393	0.197	0.203	0.823
Standard dev.	0.274	4.3	0.712	0.024	1.677	0.041	0.039	0.154	0.031
Minimum	0.366	8.8	11.3	0.412	-0.916	0.271	0.102	0.023	0.726
Maximum	2.797	30.1	15.6	0.5	8.397	0.51	0.29	0.509	0.869

6. For example, on the 1984 initiative to prohibit public funding of abortions, the mean county level vote in favor of the liberal position (not prohibiting public funding) is 45.54%. The mean vote for the Democratic voting variable is 40.3%. In Denver county, 58.4% favored the liberal position on abortion, and 57.1% voted Democratic. The dependent variable is calculated as:

$$\frac{58.4/45.54}{57.1/40.3} = .901.$$

In Denver county, voting on social issues and party voting are fairly consistent, with the county being somewhat less liberal on the abortion issue than party voting would indicate. For the same initiative, Arapahoe county provides an example of social proactive voting. Here, 52.3% favored the liberal position on the abortion issue, but Democratic voting is only 38.1%. Using the method of calculation described above, this county scores 1.22 on the dependent variable. Castilla county exhibits social reactive voting on this issue. Only 36.9% favored the liberal position on abortion, while 63.5% voted Democratic. Castilla county scores .52 on the dependent variable.

APPENDIX B: Aggregate OLS Models: Average Score on Proactive Voting for Colorado Counties on Eight Ballot Initiatives

	1	2	3
Constant	2.038*** (.401)	3.32*** (.539)	2.697*** (.676)
Control variables			
Party competitiveness	-5.153*** (.709)	-5.593*** (.659)	-6.072*** (.728)
Population density (logged)	-0.006 (.011)	-0.028*** (.009)	-0.027** (.011)
Median income (thousands)	0.011** (.005)	—	0.002 (.006)
Median education	0.094** (.031)	—	0.055 (.037)
Structural variables			
Income inequality	—	-0.292 (.319)	-0.485 (.354)
Educational inequality	—	-5.031*** (.551)	-4.003*** (.929)
Industrial heterogeneity	—	1.615*** (.448)	1.595*** (.457)
Ethnic heterogeneity	—	0.769*** (.148)	0.749*** (.149)
R^2	0.593	0.744	0.755
Adjusted R^2	0.565	0.717	0.719
(n=63)			

[a] Standard errors are in parenthesis.

Joint significance of structural variables

	1	2	3
	—	F = 23.68	F = 10.41
	—	df = 4/56	df = 4/54
	—	$p < .01$	$p < .01$

* $p < .10$ ** $p < .05$ *** $p < .01$

This variable takes on close to a normal distribution with a slight positive skew. The mean is 1.016 with a standard deviation of .274. The median value is .990.

7. As is always the case, the variables used in this analysis do not exhaust the supply of potentially interesting or informative variables. I have chosen variables that capture the extent of differentiation on basic parameters such as ethnicity, occupation, income, and education, that have been shown in the past salient in the formation of political interests. An obvious omission is a variable measuring religious heterogeneity. Undoubtedly, religion plays a role in many of the social issues included in this study. The omission of a religion variable is not necessarily fatal, however, since religiosity and religious preference are associated, to a great extent, with variables reflecting ethnicity, education, income, and occupation. One might conceive of religion as being an intervening variable in this case. Nevertheless, if county level data on religiosity or religious preference were available, it would be a valuable addition.

8. I also employed a random effects model. The results (not shown) were not substantially different from the results of the fixed effects model.

9. The issue specific dummy variables, also do not seem to have an impact on the model. A test of the joint significance of the issue specific dummy variables, in all three models, yields an *F* value that is far from having statistical significance. Tests of the joint significance of the four structural variables, however, are significant.

10. As a check on these findings I also run the analysis without utilizing a panel design. Instead, I simply take the counties' mean score on the dependent variable across the eight initiatives and regress that on the independent variables. In this approach, the signs of the coefficients are the same as in the pooled model. The sizes of the coefficients are also very close to those of the pooled model. Income inequality is not statistically significant, while the coefficients for educational inequality, ethnic heterogeneity, and industrial heterogeneity are highly significant and in the expected directions. Since the effects of the independent variables are much stronger on some of the issues than on others, aggregating the dependent variable, as opposed to using a pooled model, substantially increases the R^2. Nearly 76% of the variation in proactive voting is explained by the full model. The results of the aggregate models are reported in Appendix B.

11. Each structural variable is added one at a time, beginning with ethnic heterogeneity, and then followed by income inequality and industrial heterogeneity. I add educational inequality last since it is highly correlated with median education and with ethnic heterogeneity. It is interesting to note that in almost every instance, the addition of an inequality variable increases the coefficients of the heterogeneity variables in the model. In the same manner, the addition of a heterogeneity variable reduces the coefficient for the inequality variable. Ethnic heterogeneity enters the basic model negative and statistically significant on the 1980 lottery vote, and negative and of borderline significance for the 1988 abortion vote. In the first case the coefficient becomes positive and statistically insignificant when educational inequality is added. For the 1988 abortion vote, ethnic heterogeneity is no longer significant when income inequality is added. For the environment vote and for bilingualism, ethnic heterogeneity enters the model positive and highly significant and remains so with the addition of each structural variable. For the alcohol vote, ethnic heterogeneity becomes and remains positive and significant when income inequality is added to the model. For the 1988 gambling vote, ethnic heterogeneity only becomes positive when educational inequality is added. It is also interesting that when educational inequality is not included in the model, the partial coefficient for income inequality has at least borderline significance for the 1980 lottery vote, the 1982 alcohol vote, the 1984 abortion vote, the 1988 abortion vote, and 1990 gambling vote. When educational inequality is included, income inequality is significant only for the 1984 abortion vote and the 1990 gambling vote.

References

Abramson, Harold. 1973. *Ethnic Diversity in Catholic America.* Wiley.

Alba, Richard. 1976. "Social Assimilation among American Catholic National-Origin Groups." *American Sociological Review* 41:1030-46.

Amenta, Edwin, and Yvonne Zylan. 1991. "It Happened Here: Political Opportunity, the New Institutionalism, and the Townsend Movement." *American Sociological Review* 56:250-65.

Barbosa, Luiz. 1993. "The 'Greening' of the Ecopolitics of the World System: Amazonia and Changes in the Ecopolotics of Brazil." *Journal of Political and Military Sociology* 21:107-34.

Beisel, Nicola. 1990. "Class, Culture, and Campaigns against Vice in Three American Cities." *American Sociological Review* 55:44-62.

Blau, Peter. 1977. *Inequality and Heterogeneity: A Primitive Theory of Social Structure.* Free Press.

Blau, Peter, and Joseph Schwartz. 1984. *Crosscutting Social Circles: Testing a Macrostructural Theory of Intergroup Relations.* Academic Press.

Blau, Peter, and Otis Dudley Duncan. 1967. *The American Occupational Structure.* Wiley.

Clarke, Alan. 1987. "Moran Protest, Status Defence and the Anti-Abortion Campaign." *British Journal of Sociology* 38:235-53.

Clemens, Elisabeth. 1993. "Organizational Repertoires and Institutional Change: Women's Groups and the Transformation of U.S. Politics, 1890-1920." *American Journal of Sociology* 98:755-98.

Cohen, Jean. 1985. "Strategy or Identity: New Theoretical Paradigms and Contemporary Social Movements." *Social Research* 52:663-716.

Eckberg, Douglas. 1988. "The Physicians' Anti-Abortion Campaign and the Social Bases of Moral Reform Participation." *Social Forces* 67:378-97.

Gamson, William. 1975. *The Strategy of Social Protest.* Dorsey Press.

Gusfield, Joseph. 1963. *Symbolic Crusade: Status Politics and the American Temperance Movement.* University of Illinois Press.

Hahn, Harlan, and Sheldon Kamieniecki. 1987. *Referendum Voting: Social Status and Policy Preferences.* Greenwood Press.

Heer, David. 1974. "The Prevalence of Black-White Marriage in the United States." *Journal of Marriage and Family* 36:246-58.

Hollingshead, August. 1950. "Cultural Factors in the Selection of Marriage Mates." *American Sociological Review* 15:619-27.

Inglehart, Ronald. 1987. "Value Change in Industrial Societies." *American Political Science Review* 81:1289-1303.

Jenkins, Craig. 1985. *The Politics of Insurgency: The Farm Workers Movement in the 1960s.* Columbia University Press.

Jenkins, Craig, and Charles Perrow. 1977. "Insurgency of the Powerless: Farm Workers Movements (1946-1972)." *American Sociological Review* 42:249-68.

Klandermans, Bert. 1984. "Mobilization and Participation: Social-Psychological Expansions of Resource Mobilization Theory." *American Sociological Review* 49:583-600.

Laczko, Leslie. 1987. "Perceived Communal Inequalities in Quebec: A Multidimensional Analysis." *Canadian Journal of Sociology* 12:83-110.

Leahy, Peter, and Allan Mazur. 1978. "A Comparison of Movements Opposed to Nuclear Power, Fluoridation and Abortion." *Research in Social Movements, Conflict and Change* 1:143-54.

Lipset, Seymour Martin. 1981. *Political Man: The Social Bases of Politics.* Johns Hopkins University Press.

Luker, Kristin. 1984. *Abortion and the Politics of Motherhood.* University of California Press.

Marwell, Gerald, and Pamela Oliver. 1993. *The Critical Mass in Collective Action*. Cambridge University Press.

McAdam, Douglas. 1982. *Political Process and the Black Protest Movement 1948-1970*. University of Chicago Press.

McCarthy, John, and Mayer Zald. 1977. "Resource Mobilization and Social Movements: A Partial Theory." *American Journal of Sociology* 82:1212-41.

Mitchell, Robert. 1979. "National Environmental Lobbies and the Apparent Illogic of Collective Action." Pp. 87-121 in *Collective Decision Making*, edited by C. Russell. Johns Hopkins University Press.

Modavi, Neghin. 1991. "Environmentalism, State, and Economy in the United States." *Research in Social Movements, Conflict and Change* 13:261-73.

Nielsen, Francois. 1980. "The Flemish Movement in Belgium after World War II: A Dynamic Analysis." *American Sociological Review* 45:76-94.

____. 1986. "Structural Conduciveness and Ethnic Mobilization: The Flemish Movement in Belgium." Pp. 173-98 in *Competitive Ethnic Relations*, edited by Susan Olzak and Joane Nagel. Academic Press.

Nielsen, Francois, and Art Alderson. 1993. "Income Inequality and the Great U-turn: Evidence From North Carolina Counties 1970-1990." Paper presented at the 1993 meetings of the American Sociological Association in Miami, Florida.

Nygard, Frederick, and Arne Sandstrom. 1981. *Measuring Income Inequality*. Almquist & Wiskell.

Oberschall, Anthony. 1973. *Social Conflict and Social Movements*. Prentice-Hall.

____. 1993. *Social Movements*. Transaction Books.

Offe, Claus. 1985. "New Social Movements: Challenging the Boundaries of Institutional Politics." *Social Research* 52:817-68.

____. 1990. "Reflections on the Institutional Self-transformation of Movement Politics: A Tentative Stage Model." Pp. 232-50 in *Challenging the Political Order*, edited by Russel Dalton and Manfred Kuechler. Oxford University Press.

Olzak, Susan. 1982. "Ethnic Mobilization in Quebec." *Ethnic and Racial Studies* 5:253-75.

Pinard, Maurice. 1975. *The Rise of a Third Party: A Study in Crisis Politics*. McGill-Queen's University Press.

Piven, Frances Fox, and Richard Cloward. 1979. *Poor People's Movements: Why They Succeed, How They Fail*. Vintage Books.

Schattschneider, E.E. 1960. *The Semisovereign People: A Realist's View of Democracy in America*. Hold, Rinehart & Winston.

Shefter, Michael. 1983. "Regional Receptivity to Reform: The Legacy of the Progressive Era." *Political Science Quarterly* 98:459-83.

Simmel, Georg. 1950. *The Sociology of Georg Simmel*. Free Press.

Skocpol, Theda. 1992. *Protecting Soldiers and Mothers: The Political Origins of Social Policy in the United States*. Belknap Press.

Staggenborg, Suzanne. 1988. "The Consequences of Professionalization and Formalization in the Pro-Choice Movement." *American Sociological Review* 53:585-605.

____. 1989. "Organizational and Environmental Influences on the Development of the Pro-Choice Movement." *Social Forces* 68:204-40.

Tilly, Charles. 1978. *From Mobilization to Revolution*. Wesley Publishing Company.

Unnithan, N. Prabha, and Jay Corzine. 1987. "An Application of Status Politics Theory: Alcohol Prohibition in India and the United States." *Comparative Social Research* 10:185-204.

Weakliem, David. 1991. "The Two Lefts? Occupation and Party Choice in France, Italy, and the Netherlands." *American Journal of Sociology* 96:1327-61.

Wood, Michael, and Michael Hughes. 1984. "The Moral Basis of Moral Reform: Status Discontent vs. Culture and Socialization as Explanations of Anti-Pornography Social Movement Adherence." *American Sociological Review* 49:86-99.

Insult and Inclusion: The Term *Fag Hag* and Gay Male "Community"*

DAWNE MOON, *University of Chicago*

Abstract

The term fag hag *is normally used in gay male culture to describe a straight woman who associates with gay men. This article uses intensive interviews with gay and bisexual men to explore what the term indicates about contradictions emerging from dominant views of gender and sexual identity. Other sociological studies provide explanations for why women and gay men form relationships and how these relationships are negotiated. This article explores what the term, residing at the nexus of discourses of sexuality and gender, tells us about (1) how dominant, heterosexual culture's assumptions threaten gay identity and culture; (2) the tensions present within gay male culture's own dominant discourses, tensions which certain straight women create and/or highlight; and (3) the limitations imposed on political mobilization by a discourse that posits the existence of a coherent "gay community."*

fag-hag: heterosexual woman extensively in the company of gay men. Fag-hags fall into no single category: some are plain Janes who prefer the honest affection of homoerotic boy friends; others are on a determined crusade to show gay boys that normal coitus is not to be overlooked. A few are simply women in love with homosexual men; others determine to their chagrin that their male friends are charming but not interested sexually. No matter how you cut it, fag-hag has an ugly ring to it. Syn: faggot-chaser (*pej*). *The Queen's Vernacular: A Gay Lexicon* (Rodgers 1972).

Interviewer: What are fag hags?
Billy:[1] Women, straight women, that hang out and run around with gay men. Some of whom I've known to sexually seduce gay men. That "one-good-woman" concept. . . . Also . . . one of my sisters used to run around with us and hang out. 'Cause we were fun, we did good drugs, we always danced, you know, and, but I never considered her a fag hag. She was not *interested* in dating any of my friends, so that's why I make the distinction.

* *Special thanks are due to Andrew Abbott, Julie Brines, George Steinmetz, and the reviewers of this journal, whose comments have been most helpful, and to the numerous friends and colleagues who assisted me with various versions of this article. Most importantly, I give thanks to the many respondents whose insights were both fascinating and crucial. Direct correspondence to Dawne Moon, the Department of Sociology, 1126 East 59th St., University of Chicago, Chicago, IL 60637.*

 Social Forces, December 1995, 74(2):487-510

Interviewer: What do you think makes someone a fag hag?
Henry: [They are] women who choose to participate with homosexual men, not lesbians so much, because they relate best to them, they don't relate well to other women maybe, and not to straight men either, it's just, that's the group they relate best to.

Darren: "What *makes* them be that way?" this is a topic from, for you know, for the Montel Williams show, or . . . "What *makes* women want to *associate* with gay men?" I think an open-minded liberal attitude, so not being hung up, on sexuality to begin with, um, having met and had a positive relationship with someone who is gay or lesbian, so that you are not afraid that they're abnormal or perverse people and they're just normal people.

Interviewer: What does it mean, to the person referred to, to be called a fag hag?
Dean: It's a derogatory statement, in my experience, anyway. It means a masculine bitch, with a very big mouth, who will flaunt her relationships in public by acting like "one of the guys," making "cock-sucking" remarks or whatever — pardon the expression, and by bringing up your personal life whether you want her to or not. It connotes insincerity on their end of the friendship, and yours.

The term *fag hag* comes from gay male culture in the U.S.; while participants in this culture differ in opinion on to whom exactly the term applies and when, if ever, it is appropriate, it clearly denotes a woman, usually a straight woman, who associates either exclusively or mostly with gay men.[2] Gay men may use the term to refer to women who are friends and confidantes; they may see fag hags as "pathetic," as especially "together," or simply as women who happen to have gay men as friends. Traditionally, the term is an insult, yet it may be used in such a way that no offense is meant or taken. Why does gay male culture include such a term for certain women? What does this term, residing at the nexus of gender and sexuality, tell us?

I explore what leads some participants in gay male culture to use the term, and furthermore, why those who use it do so in such varied contexts. The term hints at how gay men, and other participants in their culture, perceive the women who enter gay "space," and how those perceptions change.[3] It calls attention to the strengths and limitations of "community" as a rhetorical strategy for a movement for gay/lesbian/bisexual/transgender liberation. Furthermore, it highlights areas of conflict within popular gay male definitions of "who we are," conflicts that arise from the various ways people reconcile the contradictions inherited from dominant discourses of gender and sexuality.

To learn about gay men's experiences with the term *fag hag,* I used a snowball sample and an intensive interview format covering 33 men who identified themselves as gay or bisexual and who were familiar with the term. Interviews took between a half hour and two and a half hours. Respondents ranged in age from 17 to 68, exhibiting a broad range of qualities within gay culture. They varied in the degree to which they considered themselves "out," or openly gay, and also in levels of involvement in gay cultural and political activities. Furthermore, they varied widely with regard to their opinions on the definition and propriety of the term *fag hag,* as well as in how often, if ever, they used the term.

While it is a strength of the interview method of data collection that we can get at people's experiences and use them to test both theories and popular rhetoric, "experience" itself is shaped by rhetoric and ideology, as Scott (1992) so convincingly points out. As she asserts, what rhetoric and world views predominate will indeed affect how people experience gender, sexual identity, and relations among people in different categories along these axes. Thus, as people's experiences vary, so will their criteria for judging whom they consider to *belong* in gay space. Indeed, the boundaries of gay space are a site of constant redefinition, and these definitions are influenced by the various points of tension within gay male culture. The term *fag hag* resides at the boundary between who belongs in gay space and who does not, highlighting the tensions within a discourse of gay identity, tensions related to people's experiences of community, gender, and sexuality.

By exploring these issues, we can further understand the dynamics of heterosexism and sexism and how these dynamics are supported or subverted. As I will discuss further, gay male culture consists of multiple discourses of gender and sexual identity. Participants in this culture create it in an ongoing process, and they do so coming from diverse places and eras, bringing with them a variety of experiences and understandings of their world. Thus, on different occasions different tensions emerge as salient and terms such as *fag hag* come to have a variety of meanings.

It is important to set up some working definitions of terms. I use *straight* to refer to persons who identify themselves as heterosexual. What I call *heterosexual culture* dominates most U.S. cultures; it consists of the set of ways in which people's sexuality is limited by usually unquestioned assumptions (at times backed up by outright coercion) that sexual attraction is only "natural" if it occurs, along with other limitations, within certain social groups and across genders. Because heterosexuality structures the dominant culture, constantly being made to seem both natural and universal, it could be described as a hegemonic order, retaining a position of apparently total social authority in spite of everyday resistance.[4] The term *fag hag* parodies that order, at times subverting and at times reproducing it.[5] That "fag hags" exist suggests that even though most natives of the U.S. were and are raised in heterosexual culture, not everyone who is straight leaves all its fundamental assumptions unquestioned, and not everyone who is not straight questions all these assumptions.

I use *gay* to describe men who identify themselves as such and who participate in "gay [male] culture." "Gay culture" refers to the common system of gay language, history, and pastimes shared by those who identify as such. These people may or may not consider themselves "homosexual," since bisexuals and others who do not conform to the dominant model of heterosexuality may also at times be seen as insiders.[6] As a subculture in U.S. society, gay male culture is defined from both inside and outside, in relation to and in opposition to, heterosexual culture's model of "natural" sexual and gender relations. Within that culture, however, participants bring numerous overlapping and competing understandings of gender and sexuality. The culture is always affected by dominant discourses, and as different understandings of what it is to be gay develop, tensions emerge within gay male culture. These

tensions are at times highlighted and at times created by both the term *fag hag* and by those to whom it refers.

In the U.S., gay men and lesbians have determined continually the boundaries of their space vis-à-vis the dominant culture in order to communicate among themselves while protecting themselves and each other from persecution (e.g., Chauncey 1994; D'Emilio 1983; Goodwin 1989; Kennedy & Davis 1993). In recent decades, however, much weight has been given to the concept of "the gay community." Among both insiders and outsiders, gay community is a notion often used to describe a politically mobilizable, clearly defined group formed around members' perceived shared interests. These interests are supposed to emerge from members' common experience of being gay.

As I will discuss further below, the term *gay community* is ambiguous and problematic. Patton (1993) illustrates that the rhetoric of community and its related politics of identity evoke a sense of duty to a cause. While the term's ambiguity lends it rhetorical strength as a device for mobilizing political constituents who share certain sexual identities, it also masks existing inequalities that can lead to disunion among those it is supposed to unify. Furthermore, a notion of gay community rests on the assumption that being gay entails sharing compatible political interests with others who are gay, while positing that "gay" people's interests always oppose the interests of those who are not gay. At the end of the article, I make it clear that the existence of straight women who associate with gay men and who participate in gay cultural pastimes poses a problem for such a conception of community.

This article starts by examining respondents' views of women and the term *fag hag*. I discuss the more traditional use of the term as an insult, placing the term and the women it represents in the context of gay culture's ongoing process of self-definition. I then move on to explore ways in which the term is interpreted not as pejorative, and explore possible reasons that the term remains ambiguous. Finally, I examine what the existence of fag hags suggests for students of political mobilization, specifically for those interested in the politics of sexual liberation.

Tensions within Gay Male Culture

How a speaker conceives of or perceives "the gay community" affects how he or she defines the term *fag hag*. In fact, as a term that denotes "otherness" in the fields of both gender and sexuality, *fag hag* appears at points of tension within the ongoing discourse of what *being* a gay man means to gay men themselves. As with other identities, people sharing a gay identity look for things they have in common, yet sometimes they reach different answers to those questions.[7] In general, men who ally themselves with gay culture see themselves as in some way opposing heterosexist domination. What exactly that rejection entails, however, and therefore what it *means* to *be* a gay man, vary as widely as people's experiences of gender and sexuality vary.

People's experiences and understandings of sex and gender vary, and furthermore, as I mention earlier, their personal experiences are both shaped

and understood through varying rhetorics and ideologies of sex and gender. Because of these differences, points of contestation emerge within gay cultures as to what it means to be gay and to reject heterosexist domination. These points of contestation include: (1) Given that heterosexist ideology would have all men love only women, can a gay man resist heterosexism and still love (or even like) women? (2) Given that the dominant culture often seeks to limit people's sexual freedom by prescribing when, how, and with whom an individual should have sex, does rejecting that dominant culture mean that gay men should be limited from having "heterosexual" relations? (3) Although people retain certain privileges when they do not reject heterosexuality, should gay men get these privileges despite opposing heterosexist domination, or should they reject the very existence of these privileges? How these questions are resolved at a given moment will be reflected in who is perceived as an outsider and how that outsider is treated. Therefore, *fag hag,* taken to denote "a woman who associates with gay men," varies in connotation depending on what it means to be a woman among gay men in different situations.

As I examine uses of *fag hag* among respondents, I consider salient issues within gay culture. I begin by looking at some examples of past uses of the term *fag hag,* suggesting where gay discourse derives some of today's major themes. I then dwell a bit longer on what the term can mean contemporarily. I take observations of uses of the term from respondents who participated in gay cultures from the 1940s through the 1970s. People who came of age in the same period share definitions of *fag hag,* and the interviews suggest that certain issues dominated at these times and shaped the dominant perceptions and issues that persist to this day.

Of the four respondents older than 50 that I interviewed, three were involved in gay culture to some extent through the 1940s, 1950s, and/or 1960s. In this period, the meaning of being gay changed greatly, but it was a period in which gay culture could overall be described as underground, where language and pastimes developed as ways for people to protect themselves and each other from persecution. Thus, "the gay world" came into contact with few outsiders of any kind (e.g., Bérubé 1990; Chauncey 1994; D'Emilio 1983; Kennedy & Davis 1993).[8] Fag hags existed, however, and opinions of them reflected the tensions prominent at the time.

Respondents whose participation in gay culture began in the 1940s or early 1950s define fag hags as "lonely" women who, for whatever reason, are outsiders to the heterosexual world.

Interviewer: What do you think makes someone a fag hag?
Lyndon: Some loneliness, a need to be accepted, and if a gay group accepts her, that's important. . . . I conceive a fag hag as a woman who enjoys hanging around gay people to the extent that she almost ignores heterosexual society. A woman who is lonely, who has no boyfriend but would like one and who uses gay society as a substitute.

Interviewer: Is any woman that associates primarily with gay men a fag hag?
Karl: I suppose not necessarily, but one who doesn't seem to have much interest in straight men, who may have some unresolved sexual conflicts in her life. Anyway she gets more attention from the gay men than she would from the straights; she seems to prefer their company. I suppose theoretically there could be somebody out there, in fact I know a woman to whom I'm out to. And she has lots of gay friends but she's not, I

wouldn't call her a fag hag in any sense of the word because she also has many nongay friends and she doesn't derive all of her strokes or social benefits from associating with gay men.

Like younger respondents, older respondents see the term as referring to "fat" and/or "unattractive" women, who stereotypically "can't get a date" — women who somehow *fail* to perpetuate the sexist, heterosexual ideal. These women may be "friends," "confidantes," and "mother figures," but they are also often seen as women who, like gay men, have been rejected from or have rejected the heterosexual mainstream. Indeed, in some instances, mainstream culture's rejection of these women is mirrored by this particular use of *fag hag*. This view carries over into today, as is clear from these respondents' opinions of the term's meaning and of the people.

In the 1970s, after Stonewall, gay men and lesbians made their cultures more visible to the mainstream than they had been previously, while focusing their political aims toward sexual liberation. During this time, even as some of the attitudes of earlier decades no doubt persisted, new issues emerged as well. One example has to do with the very visibility that the gay liberation movement was trying to secure. While coming out of the closet, or making gay identities visible, was a central focus of the gay liberation movement, such visibility also encouraged heterosexual "tourism" into gay venues. As Billy remarks:

Um, I would guess [the term] grew out of the, as the gay community grew as a bigger and bigger subculture, before the late '60s when people really started coming above ground as a gay community. I think it was one of those derogatory terms that set people apart, you know, and it may have evolved due to infiltrators. Infiltrators being straight folks that hang around in gay bars because they're perceived to be more fun. Um, it wasn't as big a deal to me in the '80s as it was in the '70s, but in the '70s I would be pissed, if a straight couple came into a gay bar, I, I would want them to leave. Because I would, I always had that feeling of "you have the whole world, I have this crummy bar, get out!"

At the same time however, this respondent discusses a woman who associated with him and his friends in the mid-1970s, saying she was *not* a fag hag:

I also knew, one of my sisters used to run around with us and hang out. 'Cause we were fun, we did good drugs, we always danced, you know, and, but I never considered her a fag hag. She was not *interested* in dating any of my friends, so that's why I make the distinction.

Harold, who came out later in the 1970s, remarks:

Well, I think ["fag hag"] is a rather unflattering term but one does have to wonder, if this person's entire social life is hanging out with gay men, um, it's as if they don't want a personal life or something, because they're certainly not going to end up in a sexual relationship with any of the men, oh, I would not be surprised to know that some would be interested in the possibility of it, but, I wish I had known some well so I could get a better sense of how to answer these questions. I'm kind of making guesses. Well, I want to emphasize that to me it doesn't sound like a very flattering term, and I guess I don't understand why a fag hag would want to hang around with gay men.

Indeed, a good number of respondents who were active in gay culture in the 1970s refer to "fag hags" as women who want to date gay men or who are "*too* interested" in gay men, women who sexualize their relationships with gay men.

These respondents' comments highlight two themes that seem to have emerged in the 1970s and that have carried through to today. On the one hand, as Billy's description makes clear, is the concern that a fag hag does not belong; she is like the straight "infiltrators" he describes. On the other hand, both Billy and Harold, like numerous other respondents, define *fag hags* as women who want to *date* gay men, therefore rejecting the men's identities as gay. A few respondents see this as the "one good woman" concept, where a woman thinks she can "convert" a gay man. It would not be surprising if the threat of "conversion" became salient in response to the gender and sexual liberation discourse of the 1970s, which stressed social constructionism and suggested that sexuality was not biologically predetermined, but constructed and fluid. Both of these tendencies emerge in the sample when it is considered as a whole.

Having briefly periodized some themes within gay male culture, I now turn to the sample as a whole and look at what it says about respondents' attitudes about the term *fag hag* today. The themes described above carry on, and contribute to, debates within discourses on gender, sexual identity, and community within gay male culture. I first look at tensions highlighted when the term is used in its traditional way, as an insult. Then I look at contexts in which the term is used in an inverted, joking, or neutral way among close friends.

Tensions in Gay Culture and Insulting Uses of *Fag Hag*

REJECTION OF WOMEN VERSUS REJECTION OF HETEROSEXISM

As suggested in the introduction, gay men may perceive themselves at odds with heterosexual culture, possibly even seeing themselves as rejected by it. Several respondents comment that using the term *fag hag* relates to some gay men's "self-hatred" or "insecurity" about being gay. These problems may be expected, as people who defy heterosexual norms are often categorized as "sick" or "unnatural" and face threats of violence, discrimination, and ostracism. The discourses of heterosexism and sexism may put a gay man at a certain conflict, that between (1) a gay-affirming rejection of the heterosexist structure that would have all men love women, and (2) a sexist rejection of women.[9]

At some moments, gay men see a woman's gender as setting her apart from her gay male surroundings. There are instances where gay men's rejection of heterosexual culture can verge on rejecting *women*, perpetuating the dominant culture's sexist attitudes, even while such men share women's oppression under conventional heterosexual ideology. Some men may not disguise their sexism with an excuse of rejecting heterosexism. As two respondents comment:

Interviewer: Are some people more likely than other people to call someone a "fag hag"?
Harold: My guess is that gay men who aren't overly fond of women probably would be more likely to use that term.

Interviewer: Are some people more likely than other people to use the term?

Karl: I'm really not sure, you know, I, I could imagine there would be but I couldn't identify them. I suppose it shows a certain degree of hostility. I suppose some people who are more hostile towards women might be more inclined to use that term, more than others.

Another respondent attributes such negative feelings towards women to some gay men's own self-hatred, as Terry says:

it's used, in my opinion, as a clear expression of self-hate, that is to say, that gays who dislike themselves are projecting that dislike of themselves onto someone who is friendly to them. I think that is why I dislike this term so intensely. For me it is a reflection of a sort of confusion about identity and a confusion about the value of one's own identity.

At one level, then, men may use the term *fag hag* to insult women; a number of respondents see men who use the term in this way as "insecure" or "self-hating," feeling rejected by mainstream society on account of their homosexuality. At times, rebuking heterosexism can verge on rebuking women; at other times, it seems, the term most certainly indicates a case of misogyny where a woman is scorned both by heterosexual society and by gay men.

Yet to write off all use of the term *fag hag* as self-hating or sexist would blind us to many of the intricacies of gay male cultures. At times, men who do not consider themselves self-hating use the term to sanction certain women who do not seem to support gay culture's norms or the gay identities of those around them. At other times, the term will sanction certain men who seem to blur the boundaries of gay space. And at other times, the term will be used symbolically to include women as part of a group of gay men. I discuss the term's inclusive uses below, but first I examine uses of the term *fag hag* as an insult that need not betray direct sexism on the part of the speaker.

SEXUAL FREEDOM VERSUS GAY COMMUNITY

There is a nonmisogynist way in which being a woman constitutes a threat to gay male identity. This threat comes in part from a dialogue between two discourses: (1) a discourse of sexuality as a malleable construct, and (2) a discourse of gay community building as a route to political empowerment. Does challenging heterosexism imply that gay men must be forbidden from having "heterosexual" encounters? What "duties" does identifying oneself as gay include? Before addressing the possible conflict between two discourses, it is important to consider how some respondents see fag hags as a threat to gay male identity and to a sense of gay male community.

To many respondents, the most obvious use of *fag hag* describes women whose overt behavior makes them appear not to accept their gay male companions' sexuality. Several respondents see these women as a "problem." These women actively draw notice to the problem inherent in a discourse of sexual malleability. For instance, as noted above, fag hags are associated by many respondents with sexual tension or unrequited love relationships. For some, these relationships define a *fag hag* and without them, one is something else, if labeled at all: a friend or acquaintance, for instance. In cases where overt behavior elicits use of *fag hag*, the term is used as a sanctioning device to reject women who may represent heterosexist views of gender and sexuality.

Respondents recall a number of threatening behaviors that women have used around them. Fred recalls a woman who went to the extent of proposing marriage to him, and Oliver discusses a woman he and his bar friends associated with, who would habitually drink too much and proclaim her love for one of the men of the group, crying and saying "But don't you understand? I really love you." Numerous others mention women "falling in love," "wanting to date," or "promot[ing] sexual tension with gay men." Furthermore, to call such a woman a "fag hag" may mean that she vicariously "lives off of" her gay male friends, or has "no life of her own."

A good number of respondents remark upon the sadness of these situations, but a number make it apparent — whether with descriptions such as "inappropriate" or "setting themselves up to fail," or with the love object's ensuing response of "get off me!" — that these situations are awkward for both parties in question. It is difficult not only for the woman for whom "all the good guys are gay," but for the man as well, since these situations manifest the woman's lack of acceptance of the man's self-identification as gay. To these respondents, a fag hag does not understand or accept what it means to be gay; these women refuse to understand the respondents' belief that a gay man could never have any sexual or love interest in a woman.

Heterosexist common sense dictates that people may not love or have sex with whomever they want, but does challenging such domination allow for the possibility that a gay man might sleep with or fall in love with a woman? For some respondents, *fag hag* seems to refer to a woman who encourages such a transgression of the gay/straight binarism. However, some find even more threatening the possibility that a "gay" man *could* have such an interest in a woman. This threat resides at the nexus of tension between the ideal of sexual freedom and that of gay community. In the community model of gay liberation, it is often implied that gay men have a duty to act as unambiguously gay as possible (Rubin [1984] 1993). While gay community building rests on the construction of stable gay identities, the pursuit of sexual liberation is founded on a rejection of such stable identity formation.[10]

Some respondents, apparently influenced by the community-building logic, see the women they consider to be fag hags as threatening in their apparent disrespect for the norms of gay male culture; others, caught in the irony between the two logics, do not consider the term to apply only to a certain kind of women, but to any woman in gay space. That any women among gay men may be considered conspicuous or threatening calls attention to a problem emergent from the contemporary discourse of "gay community" — even if it strives for a goal of individuals' sexual freedom, it risks perpetuating the logic that there are "naturally" two complementary sexes, a logic that can neither explain nor accept experiences that escape this divide.

Women's threat within gay culture comes from this dominant binary logic as it continues into gay male discourse. The threat women can pose becomes obvious when we reflect that straight or "straight-acting" men who participate in gay culture do not have any widely used special label. One of the respondents discusses this disparity and suggests that "straight" men who associate with gay men are more likely to be considered "closeted," gay but not yet open about it. While a person may be referred to as *closeted,* this term is used less

frequently than *fag hag* as a direct insult. Since almost all gay men have to some extent identified themselves as straight before taking on a gay identity, this identity poses a less salient threat in gay culture than that posed by straight women in certain contexts.

To suggest that simply *being* a woman constitutes a threat may sound far-fetched, but the prevalent view that heterosexuality alone is natural endangers an individual's maintenance of a gay identity. When a gay man associates with a woman, even if he does not wish to hide his gay identity, many people would presume him to be straight unless he "acted" gay. This is so because heterosexual culture appears to inundate many aspects of life, from the media to the family to the legal system. From the perspective of someone who does not identify with the heterosexual mainstream, it appears that when a man and a woman are seen together the mainstream assumes that they are heterosexual; they do not challenge that dominant view of gender relations.[11] Thus, for a gay man to reject women's presence in gay space may be an attempt to challenge heterosexism, albeit one that seems short-sighted in the logic of sexual liberation.

One respondent provides an illustrative example. Steve describes a confrontational situation, in a gay bar, for instance, where one gay man approaches another who enters with a woman, snaps his fingers in the direction of the woman and asks the man, "Who's your *fag hag?!*" In such an instance, the speaker reminds the addressee that the woman is a fag hag, and therefore the addressee *is a fag*. Certainly the rudeness and sexist implications of pointedly ignoring a woman in such a situation cannot be denied, but this scenario also points to deeper issues and supports what postmodern theorists have articulated, that sexual identities are not only socially constructed but also tenuous (Butler 1993).

That the speaker may feel the need to remind a man of his supposed sexual identity suggests that the discourses of sexual identity and liberation conflict at times. In this gesture, the speaker seems to be calling attention to the potential for a sexual relationship between the fag and the hag, a situation that could be seen as betraying "the community." Indeed, by providing a model of what a man must do in order for other gay men to consider him gay, a discourse of gay identity limits acceptable behavior in ways that limit personal freedom. In this case, it limits with whom a man may associate if he wishes to go to a bar without being bothered.

If a man associates with a woman, he may look straight, but if he goes to a gay bar with her, and appears to want to be part of gay culture, he may be viewed as a marginal participant, one who has an ambiguous relation to gay culture. A participant in gay culture may call the accompanying woman a "fag hag" even if she does not actively question the value of being gay; in such a case, the term is used in reference to the woman, but to sanction the man. In a discourse of sexuality as a malleable construct, the tension arises between the desire for sexual liberation, on one hand, and the fear of gay men straying beyond the boundaries of "the gay community," potentially weakening it, on the other. This threat to the gay community's stability comes from the paradox that sexual liberation need not require a man to behave exclusively "gay," in the definition created by the dominant voice in gay culture.

HETERO-PRIVILEGE VERSUS GAY COMMUNITY: FAG HAGS AND MARGINAL MEN IN GAY SPACE

When respondents talk about fag hags in relation to marginal men in the gay scene, it becomes even clearer what kinds of symbolic threats are salient within this culture. Marginal participants in gay culture bring me to a third tension in gay discourse, this one also related to the rhetoric of gay community. Respondents see as marginal those men who appear less secure with a gay identity. These men, usually younger or "newly" gay, bring with them to the threshold of the gay world their straight female companions. Unacculturated to the ways of gay male culture, these individuals behave in ways the people around them deem inappropriate. Such behavior hits on the conflict between wanting to become a part of a community of gay people and wanting to retain the privileges that come from membership in the heterosexual mainstream. Does rejecting heterosexist domination imply taking heterosexual privileges in spite of being gay, or does it imply rejecting the very existence of these privileges?

This segment of gay life is pointed out in the data by about a quarter of the respondents, who associate the presence of fag hags with youth and insecurity about one's sexual identity. Martin points out the insecurity he ascribes to gay men who associate with fag hags:

> Others might perceive my friends as fag hags, but I don't. We don't put on the air that we're heterosexual. Sometimes in a bar, a gay man will be acting like they're in a heterosexual couple, where he'll just be spending all of his time with the woman. That's kind of a crutch, dealing with [internalized] homophobia. . . . Also some people feel uncomfortable with two men dancing together, even in a gay bar, so she's kind of a buffer. . . . I've used it negatively too. If someone is having a hard time coming out, and is playing games with people and not being honest, I'll be like, "Oh there's bla bla bla, with his fag hag, fooling the world." [If the other man is trying to pass as straight.] But that's my own self righteousness, but I feel angry at that injustice against the gay community.

The presence of a fag hag indicates insecurity on the part of the man who has her around. Martin suggests later in the interview that he has noticed it among younger gay men, of high school or college age, also perhaps indicating insecurity, if we assume people tend to be less insecure as they age. This "insecurity" differs from Terry's description cited above, where he sees men who use the term *fag hag* as indicating disbelief that a straight person would want to associate with gay men. In the case described by Martin, the man with the fag hag is not fully self-identified as gay; this weak gay identity makes it unlikely that he will participate in gay activities, and makes him wish to buffer himself from other gay men by associating with a woman.

Another respondent, Steve, describes associating with fag hags as "an adolescent phenomenon," and later continues:

> I would expect to hear it in a college dormitory, or in a nightclub . . . because it's a term used by, um, more, like, freshmen, late adolescents. . . . It's more like, a need for protection, 'cause I think it's often the relationship between the fag and the fag hag was one of self-defense. 'Cause the fag hag doesn't want to be pegged out as a single woman who doesn't have a date, and the fag doesn't want to be pointed out *as a fag*.

And Oliver explains:

She was some sort of weird — I'm thinking of [a particular woman] now — she was always some sort of strange buffer between ourselves and, that bar scene which was pretty nasty, where I was hanging out. (chuckle) . . . And I'll see it, nowadays, I'll see, and they seem like they're seventeen years old, and they got some sort of fake ID going, and they're just really silly . . . It's some sort of weird stage, for both of them.
Interviewer: What kind of stage?
Oliver: A stage in terms of, until both of them get acclimated to what sort of realm they want to function in. Like that gay man, that boy, wants to be in the gay community, and the gay bar scene in specific. This stage about getting to, where you can walk in there alone. You can't do that when you're seventeen.

These respondents, all speaking from personal experience, see a fag hag as someone a young, unacculturated gay man will cling to as a buffer between himself and the gay world. By "clinging" to a woman, the man keeps from appearing gay. If the woman is seen as the object of hostility, she will be derided as a "fag hag," as we saw with Martin's commentary at the beginning of this section. Both Steve and Oliver illustrate the relationship as one of mutual defense between the young man and the woman, where the man fears being overwhelmed by gay culture, and where the woman, for whatever reason, seems insecure, or afraid of being pegged as "a woman without a date," who fails to live by heterosexual standards. While the man may or may not be likewise derided with his own pejorative term — like *closet case* — the woman may receive the label of *fag hag*. In such cases, she is derided for her friend's insecurity. If her male friend is not fully an "insider," her own presence becomes salient and threatening. As the next section will show, if a man is an insider to gay male culture, a female friend of his may be an insider also.

As an unacculturated gay man "clings" to a straight female companion, as we have seen, he buffers himself from "the gay community." Thus, another tension within gay culture is visible in Martin's commentary. In his view ("I feel angry at that injustice against the gay community") the gay community is a collective to which a gay man has a duty to reject the comforts associated with being presumed heterosexual. This sense of duty echoes that discussed by Patton and that I discussed briefly above. Since Stonewall, this duty has been constructed through a discourse of unity and liberation. This discourse has created a conflict, however, for even as coming out has been framed as part of one's duty to *the* gay community, coming out has also been framed as a route to an individual's *personal* fulfillment. From these potentially contradictory discourses, the age-old tension between personal fulfillment and duty to community may emerge.

SUMMARY: INSULT AND NOT BELONGING

A good portion of respondents agree that it is not gender, but behavior, that makes someone deserve the label *fag hag* as a mark of opprobrium. As the above commentary from Billy suggested, a few point out that a woman gets called a "fag hag" if she is conspicuous, if she appears not to belong. This "sticking out" may be because she does not appear to accept her companions' identities as gay, or it may be because her presence as a woman challenges the

dominant gender dichotomy as it applies in gay male culture. One respondent even takes the issue of belonging a step further, remarking that even a gay man can feel like a fag hag:

Oliver: Sometimes I used to consider myself, *my* self, a fag hag.
Interviewer: What do you mean?
Oliver: In that, I used to wonder, "What am I doing here?" Um, in that I can imagine [a certain fag hag we used to hang out with in the bar] asking herself, "What am I doing here?"

In this situation, the man was gay and is still gay, so it is not his self-identity that gave rise to the experience of discomfort, or not belonging. Rather, he did not feel that he belonged in the particular group with which he associated. Like this man, a woman might "belong," and feel that she belongs, more in some settings than in others. She would be called a "fag hag" where there was a question over whether she fully accepted the gay identity and culture of those in the group with which she associated.

As almost all respondents agree that not all women who associate with gay men are fag hags, a woman's presence among gay men can be perceived as threatening or not threatening at all. In some situations, an individual's gender is hardly an issue, whereas in other situations, a woman's gender does pose a threat. While respondents' definitions of the term vary, they converge to support the idea that women, because of their *being* women, cannot be assumed to be insiders. In gay male space women "stick out," as one respondent phrases it, yet in the words of another, women cannot be judged as fag hags "based on sight calls." Unlike gay men, women are not *assumed* to be part of the gay community, but must one at a time prove themselves worthy of membership. However, as many respondents make clear, it is not uncommon for women to be fully accepted as members of a particular gay community.

Determining the Perceived Threat: Group Cohesion

Since straight women often cause and highlight tensions in gay male discourse, a woman's behavior or even perceived identity may threaten elements of gay male community, hence she will be given a negative label in some instances. I have considered some contradicting themes in gay culture's ongoing discourse of self-definition, some of the norms of gay culture and some kinds of identities and behaviors within that context that may be perceived as threatening. What I have yet to examine is what, overall, accounts for the variation that exists in determining how threatening an individual will appear, whether the threat is symbolic or overt. To answer this question, we can explore the concept of how "cohesive" is a particular group of friends, strangers, or acquaintances, to what extent individuals are likely to conform to that group's norms, and to what extent group members share tacit assumptions about their surroundings.

In one model, *group cohesion* is defined by Homans ([1961] 1974) as the extent to which members of a group understand and accept common norms, follow them, and reward one another for "good behavior." According to Homans, in any group, members value conformity to the group's norms, yet because members are more similar and exchange more rewards in a highly

cohesive group than in a less cohesive one, it is less likely that a nonconformer will be present in a highly cohesive group. He gives two reasons for this likelihood: (1) because of the "tight-knittedness" of a more cohesive group, it is less likely that individuals who do not conform the group's norms will enter it, and (2) the few nonconformers who do enter will either quickly learn to conform or else will be driven out as a result of the strength of the disapproval earned by not conforming.

Many of the individuals in the sample suggest they belong to highly cohesive groups of friends. In cohesive groups, members all understand the group's norms and reproduce them. The tensions that arise in less cohesive settings, such as those examined above, do not exist in these more tightly knit groups. In a cohesive community, because group members know, trust, and support each other, their friends' gender or sexual identity is hardly an issue, and certainly not a threat. Thus, one respondent says that the straight women in his group of friends are a "nonissue"; they do not cause problems, nor does anyone expect that they would. They are accepted as full members of the group. Homans's idea that cohesive groups have widely understood norms, and therefore more reason to conform, finds support in this case.[12]

Fag hag is also used in certain kinds of cohesive groups as well, but Homans does not help us to understand why the term could be used in such a group. In these cases, the woman being addressed does not usually pose a threat, and when the term is used, it is in an ironic or joking way. Not all members of cohesive groups use the term in this "positive" way, however, so let us look further in to this area of the term's use.

Fag Hag: Some Take It "As a Compliment, Thank You"

If a woman is not fully accepted as a member of a group of gay men, she may hear herself being called a "fag hag." If she is accepted by others in the group, she may still be called a "fag hag," odd though this may seem. In this case, where the woman does not challenge her companions' gay identity, and where she is not perceived as a threat to the politically charged reproduction of a community of gay men, she may be accepted as a member of that community. Many of the respondents would not call this woman a "fag hag," since they see it as an insult, yet some would, and would do so without meaning it derisively. In these cases, the term's meaning is inverted to symbolically include these straight women in a group where, by some standards, they would not belong. Symbolically including women in a group of gay men can often be seen as rejecting not only the gay/straight dichotomy, but also refusing to participate in a sexist rejection of women.

Linguistic symbolic inversion, or using a largely pejorative term in an ironic, and empowering way, is an interesting though little explained phenomenon. This phenomenon has been discussed largely by scholars who studied African American slang in the early 1970s. Some scholars have argued that people may invert traditional insults, such as "nigger," "bitch," or "fag," in a way that makes clear the strength of the friendship between the speaker and the hearers, as insiders know whether or not to take offense (Abrahams 1976). One author,

Brown (1972), points out that for somebody who is a black "insider" to use the word nigger makes visible that that person *belongs* in a particular group of people.

Symbolic inversion also takes place in the context of gay culture. Gay folklorist Joseph P. Goodwin (1989) discusses this phenomenon:

> [Inversion] offers a means of insulting people who are so adept at stigmatizing gays, a way of expressing contempt that frequently passes unrecognized by those who have been insulted. Inversion also conveys a defiance of heterosexual standards as well. It is a way of saying to straights, "We do not accept your morality. We have our own culture with its own ethics, and these are the rules by which we live." (15)

As Hall (1982) also says, by reappropriating one of these terms in such a way, a speaker may feel personally empowered and furthermore, may contribute to a movement of ideological empowerment. As gay culture and gay politics have focused on transforming the social value of homosexuality and a gay identity, term inversion is not an unexpected phenomenon in gay culture.

Goodwin discusses inversion of terms that are normally pejorative. This reference applies to cases where gay men call other gay men names such as "fag" or "queer" as terms of affirmation or endearment where they would otherwise be seen as disempowering. A sizable majority of respondents in the sample point out that terms like these can be understood as joking or even endearing as long as they are used by another group member. As one respondent puts it:

> *Warren:* If anything, [fag] has become sort of a jocular term.
> *Interviewer:* In, uh, what sense?
> *Warren:* That it's a way to reassert to yourself the fact that you're okay. . . . My sexuality is certainly still an issue, and I still think about it a lot, but it's something that I have come to accept, and accept at such a basic level. . . . So, and by using the term fag on myself, it's just a jocular, joking sort of — you know, somebody else can sort of use that term to make fun of you, or to hate you, but, you know what it means [to be gay, whereas those who use these names to insult you don't understand that there's nothing wrong with being gay,] and it's okay.

As Warren and others indicate, these terms can actually strengthen the sense of camaraderie among participants in gay male culture when used in an atmosphere where tacit understandings are shared and the speaker's intention is clear. In a group of the type we have been discussing, a gay man who would not feel comfortable calling himself a fag, for whatever reason, would be quite unlikely to call a friend a "fag hag." Notably, a number of respondents who discuss the positive slant of inverting terms such as *fag* among friends and loved ones discuss such inversion as something that is acceptable "within *the community.*"

One may argue that inverting terms reproduces old prejudices while failing to create new meanings for these terms, yet such an argument places too much importance on words' ability to forever contain a set meaning. Words are made up and used by people, and slang terms, which are not usually learned with formally codified definitions, have meanings that may vary even more than most terms. Judging from the variety of responses in this sample, it appears that

people's own experiences have a great deal of influence on their definitions of the term *fag hag*.

The firm link between words and ideas has been greatly challenged by linguists and sociolinguists, and the work of Rommetveit (1974) is particularly useful for our purposes. Rommetveit indicates that in situations where people best understand each other, they require the fewest words to communicate. He gives the example of a man uttering to a woman the word "pot," a situation that would be meaningless to an outside observer, yet that would make perfect sense to someone who knew that the two were married, had a teenage son, were concerned about his unusual behavior, and understood the word "pot" as slang for the drug marijuana. To that observer, as to the woman in the situation, the husband's hypothesis about the son's drug use would be perfectly clear, because so much in the situation "goes without saying."

Of course, this example involves neither the use of an insult nor its symbolic inversion, but Rommetveit's ideas still elucidate our question. To an outside observer, the utterance "You fag!" could be interpreted with its conventional insulting meaning, yet a number of respondents recall situations where offense was not intended by another gay man making the remark. Consider two gay men who understand that they are old friends, that they are both gay and consider their orientations a blessing, and that one's wanting to play a Judy Garland record at a birthday party is a stereotypically gay activity and therefore a good one. In this situation, one gay man saying "You fag!" to the other may be seen as affirming and supportive so long as the two were confident in their agreement that enacting a gay stereotype was not bad behavior for gay men. Because stereotypically gay behavior is derided by much of the dominant culture, the utterance becomes a joke shared by those "in the know" against that derision.

A number of participants in the study draw a parallel between using terms like *fag* and *queer* among gay friends and using *fag hag* with close, straight female friends. This parallel does not quite come, however, since those using the term are not fag hags and those receiving it are not gay men. If it were two women in this situation calling each other fag hags in an endearing or ironic way, in a gesture of flouting their shared stigma, there would be a parallel, but this would be a different situation.[13]

More often, in situations where the term *fag hag* is used in an endearing way among friends, it appears that it is used by men symbolically to establish or affirm a woman's membership in the group of friends. If a gay man does not feel threatened by these women, he may view participation in gay pastimes and acceptance of a gay identity as more important than being a gay man per se. By using the term in such a way, a man may pull a female friend closer into the realm of insiders. Thus, as one respondent says, using the term can be seen as a gesture of affirmation, a way of saying, "You may not be gay, but you're a fag hag, so you're cool now." Another says calling a woman a "fag hag" is "kind of inviting a woman to share the stigma of being gay in our society." In such a use of the term, respondents make it clear that they see gender as a "nonissue," implying that they reject the sexism traditionally associated with the term.[14]

Not only can the term consciously be inverted to include women in the group, but its inclusive properties can also be less overt. A few respondents

point out that if the term is used in a situation among close friends, it is intended as a joke. It might be used in the context of making introductions, where the man making the introductions might introduce the woman with something to the effect of, "And this is our fag hag," with a wry look to convey the message, "but of course she's not really." Another joking situation that several respondents bring up is using the term to poke fun at a woman if she does something silly or slightly annoying.[15] Jokes, as Homans points out, can serve as a reward for positive social interaction. As members joke with an individual in this fashion, it seems reasonable to assume that they feel she belongs in the group. Furthermore, by sharing in such an inversion of the term, participants in the conversation share a joke against the dominant culture that would have this straight person shun gay men instead of befriending them.

People who use the term *fag hag* positively view the term as empowering for the gay community as a whole; they use it not to make an issue of the woman's behavior or sexual identity, but to reclaim the label for their own empowerment. Of course, in many instances where straight women succeed in being recognized as "members" of a group of gay men, or "gay friendly," they will not be called fag hags. In most of these cases, individuals do not see their friend's gender or sexual identity as noticeable or worthy of a label. Gender and sexual orientation are not, for these respondents, sites of tension.

Overlapping with the cases of those who do not use the term because they feel that a friend's sex and sexual identity are not problematic are those cases where the men do not feel comfortable inverting the term and using it in an ironic or joking way. One respondent says, "I am just not at the psychological stage where I can use those terms in that way." Why this would be may have to do with factors of sociological interest or may be largely due to such individuals' different backgrounds and factors beyond the scope of this study. This perspective does not appear to be uncommon, however, and should not be forgotten as contributing to the use of the term *fag hag*.

Within the arena of "positive" use of the term, we have seen that its inverted use will take place in more cohesive face-to-face groups, where the women do not pose a threat, but share in the tacit understandings of their group of friends. Gay men who, in this type of situation, would invert pejorative terms with reference to themselves, may use the term to express symbolically that these women are insiders of their community. Whether spoken or not, friends in these cohesive groups share the understanding that they resist the dominant sexual order by forming meaningful and rewarding relationships outside of the binary norms that would dictate that every sexual relationship occur between a man and a woman and that would not often allow for platonic relationships across the gender divide.

Among these respondents, symbolic inclusion is likely to be used among those who perceive manipulation of language as a route to personal and political empowerment. And although this sample is relatively small, within it those who fall into this category tended to be those more plugged into gay cultural activities and/or active in contemporary gay politics. Indeed, these men themselves could be defined as "more inside" than others. An individual seems to need a certain degree of familiarity with gay culture before he or she can play with its language. Just as one need not be a gay man to participate in gay

culture, one will not automatically be a member of that culture if one simply behaves in some way like a gay man.

One Term for Two Kinds: The Persistence of Ambiguity

As the above sections have illustrated, roughly two kinds of women involved in gay culture may be called "fag hags" — those who in some way challenge gay identity or the boundaries of a group, and those who affirm the value of being gay while challenging the limits of a strictly defined "gay culture" that would be threatened by a woman in gay space. These appear to be to radically opposed types of people indeed, yet no semantic difference exists to reflect this. It appears that the reason only one term exists, the negative, sanctioning term, is that only in a threatening context is a friend or acquaintance's gender or sexual identity salient.

The term has become more ambiguous as the value of a gay identity among nongays has become more ambiguous. Until recently, as some respondents comment, few heterosexuals had contact with groups of people who were gay or bisexual. But this situation seems to be changing. In recent decades, gay activists, commentators, and other gay leaders have stressed the importance of integration between the two sets of cultures and increased understanding (Marotta 1981). Perhaps intermingling among persons of varied orientations has blurred the boundaries that distinguish these different cultures. It has thus become more difficult for gay men to instantly discern who is a threat and who is a friend.

Because of the tenacity of gender and sexual identities, often a woman in a gay setting may be called a fag hag, as Dean puts it, "based on sight calls." Most of the men in the sample have friends who are straight women. But as Oliver's testimony suggests, even a woman who appears to accept her gay male friends' identities might one day break down and betray a lack of acceptance or understanding. Furthermore, the term *fag hag* remains ambiguous because different themes of gay culture predominate in different situations. Men with different pasts, who come out as gay or bisexual at different times in their lives, and who have different experiences with the women they do and do not consider fag hags resolve common tensions within gay culture in different ways.

A word will probably never exist to denote simply "straight female friend of a gay man," given that when a woman is threatening, her gender or sexual identity may become noteworthy. Numerous respondents make clear that they do not consider friends fag hags simply because they happen to be straight women. These women understand and accept that their friends are gay men but do not consider sexual identity as the only salient defining characteristic. Their shared understandings go beyond that level, and, regardless of outsiders' assumptions about the relationships between gay and straight people or between women and men, my respondents consider their friendships deep and meaningful.

As friends, gay men and straight women in cohesive settings may joke with each other about common outsider perceptions of their relationships. Such a joke may affirm the how close the friends are by affirming their common

understandings that have grown throughout their friendships. In addition, by using an ambiguous term the speaker may also highlight the difference between "insiders" and others. Only those sharing the speaker's underlying assumptions will understand the joke.

Because the term *fag hag* is loaded, respondents see it as a term "to be careful with." Its ambiguity will persist for quite some time. Because people interpret terms that comprise our ideas of gender and sexuality, as long as being "gay" persists as a concept similar to how it is understood today, gay men's identities will at times be threatened symbolically by being with women. Themes that emerge in one period are slow to die away, so such threats as those I examine here will undoubtedly persist. Only if gender and sexual orientation cease to be salient as structurally oppressive personal identities would such a threat disappear. What is clear from these interviews is that gay communities, and parts of the straight world as well, are broadening their boundaries as people who are gay and those who are not have come to share lasting and important relationships.

Problems for Community Mobilization

Gay political movements have used the rhetorics of identity and community to mobilize potential beneficiaries' sense of duty to their cause. However, the fact that gay men have female friends, whether or not they are called fag hags, sheds light on the problems that emerge from any strategy that seeks to build a movement based on the dominant culture's sexual categories. These problems emerge because these categories do not necessarily mirror how people view their own experiences. To the extent that these issues may intervene in the process of political mobilization, they merit attention.

First, post-Stonewall gay politics has popularized the belief that taking on a gay identity makes one belong in *the* community, and that it is the only way to belong. This dominant rhetoric of gay community reaffirms the notion that gay and straight people necessarily oppose each other. Such an opposition comes into question when we think about fag hags and their gay male friends. At times, gay men see fag hags and other women as part of their own community, even though by the dominant gay rhetoric, they would not belong in *the* gay community. If the straight-vs.-gay dichotomy finds support in a movement for sexual liberation, that movement may limit political mobilization by alienating those for whom that opposition does not mirror experience.[16] Indeed, this opposition could be said to alienate many of the people who seek alternatives to the dominant culture in the first place.

Post-Stonewall politics also bring out other issues that relate tangentially to the case of fag hags and that are largely related with the notions of belonging. The movement's heavy political investment in the existence of a unified gay community reifies the dominant culture's notion that people come in distinct sexual species with common interests. This "species view" of sexuality, critiqued by Rubin ([1984] 1993) itself creates two problems. First, based on the harmful dominant myth that species of people exist in nature and have distinct "natural" essences, this perspective undergirds a discourse that can harm people by

legitimating discrimination and violence.[17] As a number of authors point out, dominant categories of sexuality define groups that may then be persecuted directly by government and other institutions, while individuals who harm no one are often deprived of jobs, housing, their children, and personal safety by this persecution (i.e., Rubin [1984] 1993; Seidman 1993; Patton 1993). Therefore, a movement for sexual freedom does not aim only for unfettered sexual gratification, but for freedom from such persecution.[18]

Thus the dominant discourses of sexuality can have very concrete consequences that are legitimated in part by these discourses. The species discourse, for instance, serves this legitimating function — ironically as Sedgwick (1990) makes clear — since while it is based on the notion that different "kinds" of people exist in nature, it is based also on the contradictory belief that one kind is healthy and others pathological, or one kind is sacred and others sinful. Therefore, while a movement founded on a notion of sexual identity may prove useful for short-term mobilization, it seems that one based on a shared stake in sexual liberation has much larger pool of potential contributors, as it bears the potential to benefit people who do not fit cleanly into the dominant culture's categories as well as some who do.

While the rhetoric of sexual identity reifies oppressive categories, the rhetoric of community assumes that taking on a publicly gay identity gives people common interests, an assumption that proves to be false in light of the political diversity even in this small sample. Furthermore, supposing that all people who share a certain identity belong to a community masks inequalities *within* the movement's supposed constituency. The gay community is an *imagined* but nonetheless very consequential community, as Anderson (1983) characterizes nations: imagined, "as both inherently limited and sovereign." Anderson points out that the perceived "horizontality" of an imagined community at times masks existing inequalities among "members." In keeping with Anderson, Seidman (1993) points out that the notion of a gay community is invoked in spite of immense differences among its imagined constituents in terms of class, race, gender, region, cultural background, identification as gay, sexual preferences and object choice, personal style, and a host of other characteristics that often lead to inequality among supposed members. If face-to-face communities include gay men who do not behave ideally gay, other gay people in these groups may become alienated from a politics of mobilization that focuses on gay identity as a basis for community, thereby not reflecting people's experiences.[19]

At the same time, as Seidman shows, this strategy excludes potential supporters and constituents who have a stake in a politics of sexual freedom and civil rights protection. Like fag hags, other nongay people who differ from or disagree with the dominant culture's sexual norms — people who are bisexual, transgendered, asexual, or participate in S/M, for example — do not have a place in the gay/straight dichotomy. Excluding such people weakens political mobilizing by limiting the size of the political constituency as well as by reifying the harmful and erroneous gay/straight dichotomy.

Fag hags and their gay male friends live in opposition to the view that gay and straight people have radically divergent interests. These people may be alienated by a political agenda that pays too much heed to the dominant

culture's radical opposition. Yet while gay political leaders occasionally seem to publicly recognize "our straight allies," many people accept the species view of sexuality, evinced in the wonderment of many respondents and scholars that straight women associate with gay men.

In stating that female friends of gay men may be important constituents and supporters of an agenda of sexual liberation, I do not mean to imply that gay men, lesbians, bisexuals, and transgendered people should turn over leadership of their movement organizations. Indeed, it is these people outside the dominant structure that may often be best able to see its internal inconsistencies and propose progressive changes. Yet if movement organizers wish to effect lasting and important changes, supporters of an ideological movement should be mobilized not on the basis of an identity that is originally defined by the dominant discourse, but on the basis of their shared stake in liberation.

Conclusion

The term *fag hag* is located at the nexus of discourses of gender and sexuality within gay male culture. As such it also resides at a point where hierarchies of difference in gender and sexual identity may be used to define the boundaries of "gay community." The term also relates to other tensions within gay culture: those between rejecting heterosexist domination and rejecting women; between a stable gay community and individuals' sexual freedom; and between the benefits of being presumed straight and those of being seen as part of a gay community. As these tensions are resolved, they determine who is perceived as an outsider to gay culture and how such an outsider will be received.

Furthermore, the term's inclusive, inverted usage suggests that communities of gay men may be much more open to nongay non-men than the rhetoric of gay community would imply. That this rhetoric is used for political mobilization suggests that some mobilization potential may be sacrificed by the strategy of identity politics. The rhetoric of community used to mobilize gay men, for instance, does not necessarily mirror the salient relationships people experience in their own communities, as the comments of these respondents show. Indeed, any political movement for sexual liberation cannot persist for long if it excludes people based on a minoritizing and divisive view of identity, even when their experiences lead them to share antihomophobic goals and those of sexual liberation for which that movement strives.

Notes

1. To preserve confidentiality, all names used are pseudonyms.

2. Three things must be mentioned at the outset: First, for my purposes here, it seems less awkward to follow the current trend and refer to "heterosexual" people as "straight." I will do this throughout the paper, intending no value judgment to accompany the term as some people do when they use it. For my purposes, straight is simply a nonscientific term generally accepted to mean heterosexual without the connotations of an essentializing science.

Second, by using the somewhat awkward phrase, "participants in gay male culture," I wish to make clear that while "gay men" have sets of shared pastimes and beliefs, etc., one need not be gay or male to contribute to and draw from this culture. Furthermore, there are those whom one could call gay who do not share in the cultures to which I am referring. Of course, there

is not *one*, unified gay male culture, but I hope the reader will bear with me and accept that phrase as shorthand for the culture created by men with shared histories of homosexuality in the U.S. who share, among many other things, the term *fag hag*.

Third, synonyms for *fag hag* exist in gay cultures. These include the likes of *fruit fly* and *faggotina*. *Fag hag* is the most common of these terms, however.

3. This article is about the specific case of gay male culture, thus, I exclude from my analysis lesbians, bisexuals and people and groups of many other identities. To examine, for instance, the relations among gay/bisexual men and lesbians, bisexual women, drag queens, and transgendered individuals would undoubtedly complicate, in interesting and valuable ways, the views of gender and sexual identity I present here. Unfortunately, such relations are beyond the scope of this study.

Since the sample does not exclude bisexual male participants in gay male culture, I will clarify one point. Men who identify as bisexual may or may not identify with gay culture. Bisexuality has an extremely complex relationship to gay and lesbian cultures, which should not be overlooked. However, the term *fag hag* emerged in a culture of male homoeroticism, which I give the shorthand abbreviation of "gay (or gay male) culture." Thus, the culture is gay even if not all of its participants are. The relationships between women and bisexual men would add an extra degree of complexity to this analysis and would also be quite interesting to examine more closely.

4. I agree with Foucault's (1980) notion that sexuality consists of discourses that serve as vehicles for power to infiltrate the microlevel, yet towards the end of this article I make clear my agreement with those who build upon Foucault's work (such as Butler 1990, 1993; Patton 1993; Rubin [1984] 1993), suggesting ways in which discourses may be overcome by disrupting the dominant dichotomous notions of gender and sexual identity.

5. In this sense, the term *fag hag* is similar to the practice of drag in Butler's (1993) analysis.

6. As I assert later in this article, even some straight people, such as fag hags, may be seen among gay culture's "insiders."

7. A gay identity is problematic in that an actor must constantly attempt to communicate to others how he or she situates himself or herself. In this way it is similar to identity in the sense used by Stone in his 1962 "Appearance and the Self." Likewise, Patton (1993) shows that a gay identity expresses not so much about an individual's ontology as about her or his personal alliance with a movement.

Suttles (1984) also discusses identity, in his case as something people who live in the same cities use to look for common culture to create a unique and shared sense of "who we are." As all gay men in the U.S. are not united by local ties, however, there is much more room for variation in what people think it means to have a gay identity.

8. Two notable exceptions to this claim are: (1) institutional drag settings such as the stage and big city drag balls, where gay and straight people would mix to view the spectacles drag queens produced, and (2) the gay men and lesbians who embodied visible stereotypes such as that of the "fairy" and the "butch" lesbian. Indeed, while a good deal of work is being done currently suggesting that gay culture pre-Stonewall was not entirely covert, no one contends that most people who engaged in homosexual acts or even lifestyles were as visible or visible in as large numbers as they are today (Bérubé 1990; Chauncey 1994; Kennedy & Davis 1993).

9. Of course, these two stances do not necessarily oppose each other. However, as (1) male homosexuality is not in itself sexist and as (2) sexism and heterosexism are closely linked in that both enforce a system where men dominate women (at both the macrolevel and at the level of individual relations), a conflict may at times arise between resisting heterosexism and resisting sexism. The conflicts and overlap between antihomophobic and antimisogynist agendas have been discussed at length in activist and academic literature. See, for example, King's (1990) "Producing Sex, Theory and Culture" and John D'Emilio's 1983 ground breaker, *Sexual Politics, Sexual Communities: The Making of a Homosexual Minority in the United States, 1940-1970*.

10. There is a good deal of literature about gay identity politics since Stonewall, most notably D'Emilio (1983).

11. Stone (1962) makes the case that social interactions all take place in light of participants' outward appearance, and that unless another appears to be similar to one, one will not be able to identify with, or empathize with the other.

12. My findings also lend support to more recent social psychological work. For instance, Ridgeway (1982) suggests that individuals who differ in salient characteristics from others in a group will succeed in the group if they appear to be group- rather than self-motivated. That fag hags are not seen as threatening in groups where they affirm the value of being gay supports her claim.

13. Some of these women, especially cited among the younger, college-aged respondents, have begun to identify themselves as fag hags, not in any self-depricating way, but either in a way they see as "neutral" or as an empowering act of inversion. Thus, one intermediate contact, leader of his college lesbian, gay, and bisexual organization was once approached by a first-year student, a woman, at the first meeting of the year and was asked, "Can straight people be in your group? I'm a fag hag." (The answer was affirmative.) One respondent, who, unbeknownst to the interviewer had a guest at his apartment during his phone interview said, "Some of them think it's an insult, but some don't mind. For instance, the one in my apartment right now doesn't mind it. In fact, she takes it 'as a compliment, thank you.'" Background laughter made it clear that these two were sharing a lighthearted joke. An investigation of women's use of the term fag hag would make an intriguing counterpart to the present study.

14. To speculate, it seems that the inversion of the term actually contributes to a change in its meaning. For instance, one young respondent does not seem aware that the term has ever had a negative meaning. For Henry, quoted at the beginning of this paper, fag hag refers to his good friends who happen to be straight women. This respondent is twenty years old and has identified with gay culture at a time and place when women and gay men do not have much to divide them. His unawareness of the term's insulting history suggests that as the value assigned to these *relationships* changes, and as more people learn the term by hearing it used in positive ways, the meaning of the term is actually evolving.

15. "Oh, you're such a *fag hag*!" might be used in a lighthearted way, similar to, "Oh, you're such a nut!"

16. Scott's (1992) point that experience is also shaped by discourse is illustrated strikingly in the cases where people believe they could have nothing in common with those of a different sexual identity and therefore shun them, as often happens.

17. This issue is expanded upon by numerous theorists such as Rubin ([1984] 1993), Patton (1993), Seidman (1993), Butler (1990, 1993) and Sedgwick (1990, 1993), who all build upon Foucault's ideas, making the crucial point that discourses legitimate discrimination and violence. Furthermore, Patton, Butler, and others move beyond Foucault by offering a vision of overcoming this power, subverting discourse by challenging dominant categorizations.

18. I would add, following Patton, that no matter how liberating some individuals may find coming out of the closet, as long as there are stable, hegemonically defined categories of sexuality creating identifiable groups of people such as "homosexuals," people's lives will be increasingly policed by these sexual discourses, as I discuss in earlier sections. Patton makes clear that it is important not to confuse the identity rhetorics people use to interpret their place in the world with any statement of their essential being.

19. A similar debate over identity and community has occurred in the literature on Latino/as as a political constituency in the U.S. Some people are quick to point out that there are great differences among various nationalities of Latinos, and argue that the category is therefore a useless contrivance. Others point out that (1) various Latino nationalities do share political views relative to other Americans, and (2) the concept of Latino creates a greater constituency from which to draw political and economic support (Dominguez 1994). A problem for both movements is to find ways to bridge differences within a diverse constituency of people without alienating people the movements purport to represent.

References

Abrahams, Roger D. 1976. *Talking Black.* Newbury House Publishers.

Anderson, Benedict. 1983. *Imagined Communities: Reflections on the Origin and Spread of Nationalism.* Verso.

Bérubé, Allan. 1990. *Coming Out Under Fire: The History of Gay Men and Women in World War Two.* Plume.

Brown, Claude. 1972. "The Language of Soul." Pp. 134-39 in *Rappin' and Stylin' Out: Communication in Urban Black America,* edited by Thomas Kochman. University of Illinois Press.

Butler, Judith. 1990. *Gender Trouble: Feminism and the Subversion of Identity.* Routledge.

____. 1993. *Bodies that Matter: On the Discursive Limits of "Sex."* Routledge.

Chauncey, George. 1994. *Gay New York: Gender, Urban Culture, and the Makings of the Gay Male World, 1890-1940.* Basic Books.

D'Emilio, John. 1983. *Sexual Politics, Sexual Communities: The Making of a Homosexual Minority in the United States, 1940-1970.* University of Chicago Press.

Dominguez, Jorge. 1994. "Do 'Latinos' Exist?" *Contemporary Sociology* 23:354-56.

Foucault, Michel. 1980. *History of Sexuality.* Vol. 1. Vintage Books.

Goodwin, Joseph P. 1989. *More Man Than You'll Ever Be: Gay Folklore and Acculturation in Middle America.* Indiana University Press.

Hall, Stuart. 1982. "The Rediscovery of 'Ideology': Return of the Repressed in Media Studies." Pp. 56-90 in *Culture, Society and the Media,* edited by Michael Gurevitch, Tony Bennett, James Curran, and Janet Woollacott. Methuen.

Homans, George Caspar. [1961] 1974. *Social Behavior: Its Elementary Forms.* Harcourt Brace Jovanovich.

Kennedy, Elizabeth Lapovsky, and Madeline D. Davis. 1993. *Boots of Leather, Slippers of Gold: The History of a Lesbian Community.* Routledge.

King, Katie. 1990. "Producing Sex, Theory, and Culture: Gay/Straight Remappings in Contemporary Feminism." Pp. 82-101 in *Conflicts in Feminism,* edited by Marianne Hirsch and Evelyn Fox Keller. Routledge.

Marotta, Toby. 1981. *The Politics of Homosexuality.* Houghton Mifflin.

Patton, Cindy. 1993. "Tremble, Hetero Swine!" Pp. 143-77 in *Fear of a Queer Planet: Queer Politics and Social Theory,* edited by Michael Warner. University of Minnesota Press.

Ridgeway, Cecilia L. 1982. "Status in Groups: The Importance of Motivation." *American Sociological Review* 47:76-88.

Rodgers, Bruce. 1972. *The Queen's Vernacular: A Gay Lexicon.* Straight Arrow Books.

Rommetveit, Ragnar. 1974. *On Message: A Framework for the Study of Language and Communication.* Wiley.

Rubin, Gayle S. [1984] 1993. "Thinking Sex: Notes for a Radical Theory of the Politics of Sexuality." Pp. 3-44 in *The Lesbian and Gay Studies Reader,* edited by Henry Abelove, Michèle Barale, and David M. Halperin. Routledge.

Scott, Joan Wallach. 1992. "Experience." Pp. 22-40 in *Feminists Theorize the Political,* edited by Judith Butler and Joan W. Scott. Routledge.

Sedgwick, Eve Kosofsky. 1990. *Epistemology of the Closet.* University of California Press.

____. 1993. "Bringing Your Kids Up Gay." Pp. 69-81 in *Fear of a Queer Planet: Queer Politics and Social Theory,* edited by Michael Warner. University of Minnesota Press.

Seidman, Steven. 1993. "Identity and Politics in a 'Postmodern' Gay Culture: Some Historical and Conceptual Notes." Pp. 105-42 in *Fear of a Queer Planet: Queer Politics and Social Theory,* edited by Michael Warner. University of Minnesota Press.

Stone, Gregory P. 1962. "Appearance and the Self." Pp. 86-118 in *Human Behavior and Social Processes: An Interactionist Approach,* edited by Arnold M. Rose. Houghton Mifflin.

Suttles, Gerald D. 1984. "The Cumulative Texture of Local Urban Culture." *American Journal of Sociology* 90:283-304.

Why Racial Employment Inequality Is Greater in Northern Labor Markets: Regional Differences in White-Black Employment Differentials*

SAMUEL COHN, *Texas A&M University*
MARK FOSSETT, *Texas A&M University*

Abstract

This article investigates employment inequality between 25-59-year-old white and black men in U.S. metropolitan areas in 1980. Greater inequality in employment is observed in the Northeast and Midwest than in the South and West. Regression analysis shows that area variation in employment inequality is closely associated with variation in unionization strength, presence of large firms, and labor demand and that it is moderately associated with variation in race differences in age structure and education and the presence of traditionally black occupations. Regional patterns indicate that the substantially lower employment inequality in the South and West can be attributed to the more robust economies, smaller average firm size, and lower levels of unionization in the metropolitan areas in these regions. These regional differences are large and serve to offset the impact of large racial gaps in educational attainment in the South that often contribute to inequality in employment.

A pressing question in the study of racial inequality in the U.S. is why black employment remains much lower than white employment even as many other socioeconomic gaps between whites and blacks have narrowed. Since World War II, racial differences in education and occupational status have declined and income differentials have also decreased albeit by a smaller amount (Farley 1984; Farley & Allen 1987; Jencks 1991). Even residential segregation has diminished to some degree (Farley & Frey 1994; Massey & Denton 1987). In

* *Author names are listed in alphabetical order. Authorship is shared equally. We gratefully acknowledge support provided by the National Science Foundation Grant #SES 9212057 and Texas A&M University. We also thank Christopher Winship, Paula England, and two anonymous* Social Forces *reviewers for helpful comments on earlier versions of this article. Direct correspondence to Samuel Cohn, Department of Sociology, Texas A&M University, College Station, TX 77843-4351.*

contrast, the unemployment gap between white and black males has expanded rather than contracted. In 1948 the white-black employment differential was only four percentage points; by 1987, it was 10 points (Cotton 1989; Jencks 1991).

One fact that is occasionally documented but has been little emphasized and thus is not well known is that white-black inequality in employment is greater in the North and Midwest than it is in the South and West. We find in 1980 among metropolitan areas with a population of at least 250,000 and a black population of at least 25,000[1] that the median gap between white and black employment rates for noninstitutional males 25-59 years old is 15.4 percentage points in the North and Midwest but only 10.7 points in the South and West. Alternatively, employment inequality measured by the ratio of the white odds of employment to the black odds of employment has a median of 3.081 for metropolitan areas in the North and Midwest compared to a median of 2.287 for areas in the South and West.[2] Similar findings have been reported previously by Darden (1983), John Farley (1987), and Shulman (1991). Darden's analysis is hampered by a small sample size. The studies by Farley and Shulman are based on larger samples, but the authors themselves took little interest in the finding of regional differences in the white-black employment gap: Farley devoted only three sentences to this topic and Shulman discussed his regional finding in one sentence.

Regional differences in white-black employment differentials deserve greater attention. One reason is that the differences are sizable and strongly suggest that racial differences in employment opportunities are not uniform, a phenomenon that needs to be explained and interpreted. Another reason is that the pattern is counter to conventional expectations. The South, which has a history of greater racial inequality in education, occupation, and income (Fossett, Galle & Kelly 1986) and which consistently ranks lowest among regions in prevalence of attitudes favoring equality of opportunity and racial tolerance (Fossett & Kiecolt 1989) is characterized by relatively low racial inequality in employment. In contrast, the Midwest and North, which traditionally have been seen as affording greater socioeconomic opportunities to blacks, are characterized by the greatest inequality in employment. For most of this century, the Midwest and North attracted black migrants from the South because racial inequality in occupation, education, and income and measured racial prejudice have consistently been lower in these regions.

Only one other important measure of racial differentiation — residential segregation — has been documented to be higher in the North and Midwest than the South (see, e.g., Massey & Denton 1987) and evidence for this regional difference is mixed until recent decades. Massey and Denton present data compiled from a number of studies showing residential segregation to be lower in selected southern cities in 1860 and 1940 (1993:21). On the other hand, segregation measured by block data for large cities showed southern cities to have equal or higher levels of segregation than nonsouthern between 1940 and 1970 (Sørensen, Taeuber & Hollingsworth 1975). For recent decades, findings on regional differences are more consistent. Massey and Denton report lower levels of segregation in the South in 1970 and 1980 (Massey & Denton 1987) and in 1990 (Massey & Denton 1993) for analyses using tract data for metropolitan areas. Farley and Frey (1994), using similar data, also report less segregation in

the South in 1990. It is interesting to note, however, that Massey and Denton (1993) and others (e.g., Roof, van Valey & Spain 1976) have interpreted differences between residential segregation in the South and that elsewhere as anomalies that are due at least in part to the region's special historical conditions and the ecological characteristics of southern metropolitan areas. It is significant that no major statement has argued that lower levels of residential segregation in the South are indicative of greater relative socioeconomic opportunities for blacks in this region.

We address the need for more detailed analysis of regional variation in white-black employment differentials in this article. Using a regression-based decomposition of racial employment differentials in 99 metropolitan areas in 1980, we find that lower racial inequality in employment in Sun Belt metropolitan areas is primarily due to lower levels of unionization, higher rates of economic growth, and smaller firm size in these areas. These factors counteract the greater educational inequality between blacks and whites in the South that usually contributes to racial inequality in employment. We interpret our findings as being consistent with the view that racial employment differences reflect employer discrimination in hiring and that such discrimination is shaped by factors such as loose labor markets that give employers more latitude in hiring decisions and factors such as larger, more bureaucratic organizations and greater unionization, which limit managerial discretion in firing and promotion policies and tend to shift employer discrimination toward the hiring process.

Descriptive Analysis

We focus here on the gap between white and black employment rates, not the absolute rates of black employment. Low levels of black employment could well be the function of a depressed economy, which hurts blacks and whites alike. It is more important for the study of racial inequality to assess, given the jobs locally available, whether all groups have equal access to employment opportunities. We measure employment for each group relative to each group's civilian, noninstitutional population; we specifically avoid using traditional unemployment rates. The latter registers the ratio of "officially" unemployed workers to total civilian labor force. This is unsatisfactory for our purposes because adverse labor market conditions may push many otherwise willing workers out of the "official" labor force entirely. Jeremiah Cotton (1989) has shown that excluding these "nonparticipants" from official unemployment rates substantially underestimates the severity of racial gaps in employment. Thus, we focus instead on the ratio of employed civilian workers to the noninstitutional, civilian population.[3] We further restrict our focus to men in the prime working ages of 25-59 to try to minimize the potential distorting influence of area variation in school enrollment (especially college), female labor force participation, and retirement populations.

The racial inequality in employment can be measured in a number of ways. For our purposes, the most attractive measure is the relative odds of white and black employment; that is, the ratio of the white odds of employment to the black odds of employment.[4] This measure has the advantage of being simple to

TABLE 1: Relative Odds of White to Black Male Employment Rates, Age 25-59 (U.S. SMSAs 1980)[a]

Northeast (N = 18)		Midwest (N = 21)	
Philadelphia, Pa.	3.695	Grand Rapids, Mich.	4.271
Harrisburg, Pa.	3.649	Chicago, Ill.	4.115
Newark, N.J.	3.620	Milwaukee, Wisc.	3.856
Buffalo, N.Y.	3.535	Detroit, Mich.	3.824
Paterson, N.J.	3.505	Fort Wayne, Ind.	3.781
Trenton, N.J.	3.321	Indianapolis, Ind.	3.687
Pittsburgh, Pa.	3.279	Cleveland, Ohio	3.555
Rochester, N.Y.	3.187	Saint Louis, Mo.	3.530
Springfield, Mass.	2.861	Wichita, Kan.	3.372
New York, N.Y.	2.783	Youngstown, Ohio	3.269
Boston, Mass.	2.735	Omaha, Neb.	3.105
Syracuse, N.Y.	2.730	Cincinnati, Ohio	3.081
Bridgeport, Conn.	2.675	Akron, Ohio	3.036
Long Branch, N.J.	2.664	Toledo, Ohio	3.005
Albany, N.Y.	2.527	Minneapolis, Minn.	2.965
U.S. median	2.527	Dayton, Ohio	2.840
New Brunswick, N.J.	2.058	Gary, Ind.	2.818
Jersey City, N.J.	1.883	Kansas City, Mo.	2.762
Nassau, N.Y.	1.826	*U.S. median*	2.527
		Ann Arbor, Mich.	2.464
		Flint, Mich.	2.355
		Lansing, Mich.	1.894
Median	2.852	Median	3.105

[a] SMSAs containing conurbations, such as Dallas-Ft. Worth, have been shortened for space to the name of the first city, for example, Dallas.

interpret. For example, a score of 2.5 indicates that the odds of employment for whites are two and one half times as high as those for blacks. In addition, the logarithm of this ratio — the relative log-odds of employment — is appropriate for regression analysis.[5] Finally, the relative odds of employment is not sensitive to variation in aggregate employment. In contrast, the white-black difference in employment rates increases when aggregate employment falls, even when the relative odds of employment are constant.[6]

Table 1 presents relative odds of white and black employment for U.S. metropolitan areas in 1980. Unlogged odds are presented, since it is the more familiar descriptive measure. The relative log-odds used in our regression analyses appears as Appendix A. Of course, the rank ordering remains identical after the log transformation.

The estimates presented in Table 1 are computed from data reported in the 1980 U.S. Census of Population based on area-specific samples of approximately 1 in 6 persons from each metropolitan area.[7] The sample of metropolitan areas

TABLE 1: Relative Odds of White to Black Male Employment Rates, Age 25-59 (U.S. SMSAs 1980)[a] (Continued)

South (N = 45)		*West (N = 15)*	
Baltimore, Md.	3.323	Fresno, Calif.	3.772
Memphis, Tenn.	3.266	San Francisco, Calif.	3.133
Montgomery, Ala.	3.046	Portland, Ore.	2.895
Jackson, Miss.	3.003	*U.S. median*	2.527
Atlanta, Ga.	2.902	Los Angeles, Calif.	2.516
Beaumont, Miss.	2.822	Seattle, Wash.	2.326
Washington, D.C.	2.803	Phoenix, Ariz.	2.141
Richmond, Va.	2.738	Sacramento, Calif.	2.137
Birmingham, Ala.	2.736	Denver, Colo.	2.105
Macon, Ga.	2.704	San Jose, Calif.	1.975
Lexington, Ky.	2.682	Vallejo, Calif.	1.951
Baton Rouge, La.	2.660	San Diego, Calif.	1.941
Houston, Tex.	2.573	Tacoma, Wash.	1.913
New Orleans, La.	2.565	Riverside, Calif.	1.827
U.S. median	2.527	Las Vegas, Nev.	1.777
Tulsa, Okla.	2.519	Anaheim, Calif.	1.567
Dallas, Tex.	2.509		
Mobile, Ala.	2.456		
Jacksonville, Fla.	2.439		
Shreveport, La.	2.432		
Louisville, Ky.	2.404		
Columbia, Mo.	2.388		
Charlotte, N.C.	2.387		
Chattanooga, Tenn.	2.318		
Charleston, W. Va.	2.310		
Wilmington, N.C.	2.307		
Newport News, Va.	2.227		
Norfolk, Va.	2.184		
Nashville, Tenn.	2.167		
Oklahoma City, Okla.	2.153		
Little Rock, Ark.	2.149		
Greenville, S.C.	2.102		
Huntsville, Ala.	2.073		
Greensboro, N.C.	2.048		
Austin, Tex.	2.006		
San Antonio, Tex.	1.868		
W. Palm Beach, Fla.	1.804		
Knoxville, Tenn.	1.803		
Orlando, Fla.	1.737		
Ft. Lauderdale, Fla.	1.651		
Miami, Fla.	1.618		
Pensacola, Fla.	1.611		
Augusta, Ga.	1.599		
Lakeland, Fla.	1.453		
Tampa, Fla.	1.418		
Daytona Beach, Fla.	1.254		
Median	2.318	Median	2.105

[a] SMSAs containing conurbations, such as Dallas-Ft. Worth, have been shortened for space to the name of the first city, for example, Dallas.

for which the measure of employment inequality is computed includes 99 metropolitan areas with total populations of at least 250,000 and black populations of at least 25,000.[8] Given the limitations of the published tabulations, our results do not generalize to smaller metropolitan areas or to large metropolitan areas with small black populations. Data for blacks are based on the census race category "Black or Negro." Data for whites are based on the non-Hispanic white population.[9]

It is not surprising that Table 1 shows that employment inequality favors whites in all metropolitan areas in our sample. Relative odds ratios above 1.0 indicate that white odds of employment are greater than black odds of employment; the higher the value, the greater the inequality in employment. The median value for the full sample is 2.599 and indicates that in the typical city the white odds of employment are about 2.6 times greater than the black odds of employment. The lowest relative odds ratio is observed for Daytona Beach, Florida (1.254); the highest is observed for Grand Rapids, Michigan (4.271).

The data reveal considerable variation in employment inequality across areas and regions. The variation across areas is indicated by the interdecile range of 1.777 to 3.620 for the full sample.[10] Important variation in employment inequality by region is revealed by the medians for the regional subsamples. They show that racial inequality in employment is higher in the Northeast and Midwest, which have sample medians of 2.852 and 3.105, respectively, and lower in the South and West which have sample medians of 2.318 and 2.105, respectively. Other comparisons corroborate the basic pattern suggested by the regional medians. Only 12% (6) of northeastern and midwestern cities are below the median for the full sample of metropolitan areas; in contrast, fully 72% (43) of Sun Belt cities fit this description. Similarly, 18 of the 20 highest relative odds ratios (the uppermost quintile) are located in the Northeast and Midwest while 17 of the 20 lowest relative odds ratios (the lowest quintile) are located in the South and West. It is interesting that 7 of the 10 lowest relative odds ratios are located in Florida. To reiterate the most important finding, the regional differences are sharp and substantively important; white advantage in the odds of employment is consistently higher in the Northeast and Midwest.

The Case for a Decomposition Analysis

The distribution of inequality across regions documented in Table 1 is counterintuitive in light of traditional findings of marked regional differences in racial prejudice and attitudes toward equality of opportunity (e.g., Giles 1977; Pettigrew 1959; Schuman, Bobo & Steeh 1985; Sheatsley 1965; Taylor, Sheatsley & Greeley 1978) and marked regional variation in racial inequality in education and occupation (Wilcox & Roof 1978). These lead to an expectation of greater employment inequality in the South. To explain these large regional differences in racial employment inequality we adopt the following strategy. First, we identify characteristics of metropolitan labor markets that theory suggests are likely to influence racial employment inequality. We then perform a regression analysis that which estimates the effects of these variables. Then we investigate

the extent to which regional differences in the structural characteristics of the metropolitan labor markets account for variation in average levels of racial employment inequality across regions.

Our interest is thus focused on three questions. One is whether the metropolitan characteristics we identify can be confirmed as important predictors of employment inequality. Another is whether regional differences in distributions on these variables can account for regional differences in employment inequality. The final question concerns how patterns of regional distribution on the metropolitan characteristics contribute to generating regional differences in employment inequality.

In this section we selectively review potential determinants of racial employment inequality. We draw on a variety of intellectual perspectives, each of which we believe has validity in at least some analytical contexts. These perspectives suggest several variables that are likely to generate interarea variation in racial employment inequality. We do not presume that the variables developed here include all factors likely to influence racial inequality. We argue only that it includes many variables that are likely to be both important determinants of racial employment inequality and germane to explaining regional differences in racial employment inequality.[11]

HUMAN CAPITAL THEORY

Human capital theory suggests that racial differences in human capital investment will generate racial differences in employment. Some forms of human capital — namely, on-the-job experience — are acquired through employment and are thus are inextricably bound up with differential access to employment. To invoke these differences in a cross-sectional analysis is to advance an explanation that borders on tautology; blacks are more likely to be unemployed because they have less of the training that is acquired through employment. Human capital differences reflected in differences in formal education originate outside the employment process and are less troublesome methodologically.[12] In his cross-sectional analysis, John Farley (1987) found that larger racial education differences in an area were associated with larger unemployment gaps. Shulman (1987) also found that area variation in mean education was positively but differentially related to employment rates for both blacks and whites, implying that greater race differences in education will generate greater employment inequality.

The reasoning outlined above emphasizes the aggregate consequences of the individual-level effect of education on employment. An altogether different line of reasoning draws on the theory of statistical discrimination (Phelps 1972; Thurow 1975) to argue that education differences may have important consequences for racial employment inequality. This perspective does not emphasize the role of education in individual attainment per se. Rather it argues that large group differences in education may encourage employers to engage in statistical discrimination. That is, it increases the likelihood that employers will more readily rely on race as an inexpensive and readily observed surrogate for quantity or quality of education (Phelps 1972). This implies that race differences in returns to education will be diminished in cities where race differences in

education are large. Thus, the employment prospects of even better-educated blacks are reduced when lower relative educational attainments for blacks as a group induce employers to statistically discriminate against blacks.

These arguments lead to the prediction that racial education differences will have a positive effect on racial employment inequality. Confirmation of this prediction in the present context will not permit a distinction between the two theoretical motivations for expecting the relationship. However, regardless of the theoretical interpretation assigned to the effect, it is likely to have important consequences for regional variation in employment inequality since racial education differences are considerably more pronounced in the South than in other regions.

LABOR DEMAND AND QUEUING THEORY

Thurow (1975), Hodge (1973), Lieberson (1980), and Reskin and Roos (1990) have argued that employment outcomes are determined by a queuing process. In this process, employers first rank potential workers by relative attractiveness (however defined) and then hire the top-ranked job candidates and then continue hiring from successively lower ranks in the candidate queue until all available openings are filled. In this model, employers hire more attractive workers before less attractive workers and, as labor demand increases, employers must move farther down the queue and hire workers who would be excluded under conditions of slack labor demand.

Thurow's discussion of queues emphasizes the ranking of prospective employees on human capital characteristics and employer perceptions of trainability. However, any criteria including ethnic background can be used for the purpose of ranking potential workers in an employment queue (Hodge 1973; Reskin & Roos 1990). Thus, for example, in a discriminatory queue, employers may rank white applicants ahead of black applicants. Such rankings may well reflect overt prejudice as when employers choose white workers over similarly or even more highly qualified black workers. Alternatively, ethnic ranking may be rationalized in terms of presumed correlations between ethnicity and work habits, quantity and quality of human capital stocks, or other productivity-related characteristics (e.g., ethnic stereotypes and statistical discrimination). Statistical discrimination may be further encouraged by the fact that race and ethnicity are readily observable characteristics that are relatively easy to measure compared to many non-ethnic characteristics of interest to employers (e.g., quality of education, drug use, criminal records, etc.).[13] Recent work by Kirshenman and Neckerman (1991) documents the pervasiveness of prejudice-based and especially statistical discrimination by employers and strongly suggests that race continues to be an important factor in the ranking of potential employees.

When workers in labor queues are ranked at least in part on ethnic characteristics, racial employment inequality will be affected by overall labor demand. This of course is the model implicitly invoked by the familiar last-hired/first-fired scenario. Queuing theory predicts that the higher the rate of overall employment, the lower the employment gap between whites and blacks. Evidence for this relationship can be found in previous comparative research.

For example, Shulman (1987) found that the white-black employment ratio was negatively associated with total employment and Williams (1987) reported that blacks were disproportionately likely to be pushed out of the labor force altogether when global unemployment rates were high.

Regional differences in economic growth have been pronounced in recent decades with the South and West experiencing consistently stronger labor demand. In view of this, labor demand is likely to be a potent explanation of regional differentials in racial employment inequality.

ESTABLISHMENT SIZE

Dual labor market theory argues that large monopoly sector firms are more likely to discriminate than smaller firms in the competitive sector. The superior market position of large primary sector firms buffers them from having to minimize labor costs, allowing them to tolerate higher levels of discrimination than would occur elsewhere in the economy (Averitt 1968; Cohn 1985; Kalleberg, Wallace & Althauser 1981). This leads to the prediction that firm size (an available indicator of the presence of primary sector firms) will have a positive effect on racial employment inequality.

A potential problem with this application of dual sector theory is that it fails to consider predictions from an affirmative action model. This view suggest that large firms (and especially those with government contracts) may be subject to greater enforcement of equal opportunity and affirmative action guidelines. For example, Title VII enforcement is limited to firms with more than 15 employees and Executive Order 11246 rules for federal contractors applies only to firms with 50 or more employees. There is a substantial body of evidence that suggests that affirmative action programs have increased opportunities for blacks in at least some respects, although scholarly debate on this topic continues (Burman 1973; Freeman 1973; Heckman & Payner 1989; Heckman & Wolpin 1976; Vroman 1975).

One possibility, however, is that equal employment opportunity guidelines may have an effect on employment that is different from that of other factors, and more perverse. Namely, the applicability of equal employment opportunity regulations may foster more equal outcomes on wages, job assignments, promotions, layoffs, dismissals, etc. for minority workers who are employed in large firms. However, it may also lead employers and firms to be more selective in their hiring given that they may have less flexibility in their dealings with workers once they are hired. If such selectivity in hiring is correlated with firm size and if race is a factor that is salient in selective hiring, then firm size could have a positive effect on employment inequality.

We have noted arguments for two opposite and potentially offsetting effects of establishment size. However, on the whole, we expect dual labor market effects to dominate affirmative action effects. One basis for this prediction is that establishment size is highly correlated with superior market position but only weakly correlated with enforcement of affirmative action legislation. The relationship between establishment size, market share, and superior market outcomes have been well established in the empirical literature (Hodson 1978; Wallace & Kalleberg 1981). In contrast, the enforcement of affirmative action

and equal employment opportunity regulations has been sporadic and only weakly associated with any obvious industrial attributes (Heckman & Payner 1989; Jaynes & Williams 1989; Vroman 1975). In addition, equal employment regulations may promote less equality than we anticipate because of differences between their direct effects on employment inequality and their indirect effects from employer reactions to being regulated. Therefore we expect the dual sector effects to predominate.

A pattern of regional differences in firm size is not widely noted in the sociological literature. However, the existence of such regional differences combined with the predicted positive effect of firm size on employment inequality obviously could contribute to regional variation in racial employment inequality.

UNIONISM

The presence of organized labor in a metropolitan labor market may also increase employment inequality between whites and blacks. Gary Becker (1957) argued that employee prejudice may be as important as employer racism in determining racial inequality. Welch (1967) and Arrow (1972) have also advanced models in which inequality flows from white workers' resistance to working with minority workers (either due to prejudice or self-interest). Many employers claim to be reluctant to hire minorities because of perceived opposition from their preexisting and predominantly majority labor forces (Braddock & McPartland 1987). Such arguments have to be viewed with some skepticism (Reich 1981). Although organized labor has a record of racism, its racism has been concentrated in a small number of craft unions (particularly New York construction unions) that have received a disproportionate share of scholarly attention (e.g., see Foner 1974; Hill 1984; Jacobson 1968).

Ashenfelter (1971) argues convincingly that the patterns in such studies are atypical. He maintains that union discrimination will occur only when two conditions are met simultaneously. First, trade unions must have significant control over the hiring process. If employers control entry into jobs, then exclusionary policies by unions will be ineffective and irrelevant to racial inequality in employment. Second, the union must be capable of generating credible strike threats without organizing 100% of the potential work force. If the presence of unorganized workers can vitiate labor's bargaining strength, then unions have little choice but to attempt to organize all workers regardless of ethnic origin. Ashenfelter argues that these two conditions only apply to craft construction unions; he presents data that suggest that racial differences in both percentage unionized and wage rates are found primarily in construction trades (Ashenfelter 1971).

American unions, and in particular CIO unions, have been forces for reducing rather than raising racial inequality within the employed labor force. The CIO, and later the AFL-CIO, was an early and forceful supporter of the civil-rights movement (Marshall 1965). Many CIO unions were known for their militant defense of equal rights for blacks in the workplace. Among these are the Packinghouse Workers, the Electrical Workers, the United Mine Workers, and the San Francisco longshoremen (Fogel 1970; Foner 1974; Kimeldorf 1988;

Northrup 1944). There is at least some evidence that American unions overall have reduced income inequality between whites and blacks. Richard Child Hill (1974) has shown for a sample of U.S. metropolitan areas in 1960 that unionization was correlated with lower racial inequality in family income. Freeman and Medoff (1984) report findings that union membership raises blacks' wages more than whites', although the difference is statistically significant only in some periods.

These theoretical and empirical findings lead us to accept the general conclusion that union efforts are likely to reduce racial inequality among workers already employed. However, we argue that employer adaptations to organized labor may well increase racial inequality in access to employment. Collective bargaining agreements constrain the exercise of managerial discretion in firing and promotion decisions. Union contracts generally specify bureaucratic systems of job assignment and wage scales. Likewise they commonly specify grievance procedures and provide contractual protection against arbitrary dismissal. Enforcement of these negotiated rules concerning the internal labor market of the firm restrict management's capacity to discharge workers at will (Freeman & Medoff 1984; Jacoby 1985; Slichter 1960). This is likely to reduce discrimination against employed blacks by protecting them against arbitrary actions by supervisors.

At the same time, the resulting reduction in managerial discretion in firing and promotion may raise barriers to the employment of blacks. Discriminatory employers are likely to be especially reluctant to hire black workers when they know that once hired these workers will derive protection from structured labor-management relations. Thus, Jacoby (1985) argues that employers respond to externally induced limitations of their right to fire with more intensive screening of potential workers. This may lead some employers to give greater emphasis to statistical discrimination by race in hiring. More important, it may lead employers to rely to a greater degree on formal qualifications and credentials such as licenses, certificates, and educational degrees and informal factors such as employer and employee references. Screening potential employees in this way is likely to have a disproportionate negative effect on black employment.

If greater unionization does have the net effect of shifting discrimination toward the point of hire and thereby increases racial inequality in employment, we would expect unionization to have important implications for regional variation in racial employment inequality. The Northeast and Midwest have considerably greater unionization than the South (though not necessarily the West), and thus our prediction would suggest that lower employment inequality in the South may be traceable in part to its lower percentage of unionization.

OCCUPATIONAL COMPOSITION

Regional differentials in racial employment inequality may be a function of differences in occupational composition. Traditionally, blacks have been underrepresented in professional, managerial, sales, clerical, and skilled blue-collar occupations. In part this may reflect a degree of uniformity in norms regarding the "appropriateness" of minority employment in certain occupations.

Alternatively, it might reflect consumer discrimination (e.g., professional specializations serving predominantly majority clientele). This would mean that, even in the absence of differences across areas in the propensity to discriminate, variation in racial employment inequality might follow variation in the supply of "appropriate" employment opportunities. Overall, these differences in occupational composition would trace to differences in industrial composition and the technical requirements of each industry. However, the proximate effect is likely to be felt through occupational composition. Thus, we would expect to see an inverse relation between racial inequality in employment and the prevalence of occupations "favorable" to black employment. If such occupations are regionally concentrated, then this would contribute to an explanation of regional differences in employment inequality.

AGE STRUCTURE

Another possibility is that the racial employment differences reflect differences in age structure. Age groups vary considerably in their rates of participation in the labor force. Workers in the "prime" years (30-49) have the highest labor force participation rates. Younger adults and older adults have weaker attachment to the labor force. Cities do differ significantly in their age structures and white and black age structure may vary independently. We noted earlier that seven of the ten cities with the lowest relative odds are located in Florida. One possible explanation for this is that the age composition of whites in Florida cities is skewed toward older ages with lower labor force attachment. With this in mind, we predict that employment inequality will tend to be greater in areas where whites have age structures that favor greater labor force participation and employment. If racial differences in age structure vary by region, this factor may then contribute to an explanation of regional differentials in racial employment inequality.

SPATIAL MISMATCH

Race differences in access to jobs resulting from residential segregation has been suggested as a possible explanation of white-black employment inequality (Kain 1968; Kasarda 1990). To the extent that "spatial mismatch" generates employment inequality and varies by region, it could contribute to regional variation in racial employment inequality. This possibility is suggested by the fact that studies advancing the spatial mismatch hypothesis have tended to focus on large urban areas in the Northeast and Midwest.

Unfortunately, it is impossible at this time to subject the spatial mismatch hypothesis to a direct test in a comparative analysis because satisfactory measures of mismatch between spatial distributions of residence and jobs are not available for metropolitan areas. Residential segregation cannot be used as a proxy because it reflects residential distributions only and does not register the distribution of jobs (Galster 1987). Measures of mismatch based on central city-suburb distinctions are simply too crude. They reduce urban space to a dichotomy and ignore job and residential distributions within the suburban ring and the central city. This is a potentially significant problem in cities such as

Chicago where it matters very much whether a job is located on the north or south side of the city (Ihlandfeldt & Sjöquist 1989; Kain 1974). Perhaps an even more important problem is the fact that central city definitions are more expansive for newer cities in the South and West (due to the greater opportunities these cities have had to grow by annexation) and thus cross-region comparisons of mismatch measures based on central-city suburb distributions are highly questionable. In view of these problems, investigation of mismatch effects must be left to future research.[14]

RACIAL INTOLERANCE

Variation in racial inequality is sometimes explained with a simple culturalist model of discrimination (Heer 1959; Myrdal 1944; Schuman, Bobo & Steeh 1985). Presumably, white advantages in employment would be greater in areas where attitudes of white race prejudice and intolerance are most pronounced. However, this would not seem to be a likely candidate for explaining the lower racial employment inequality in the Sun Belt since the literature on white racial attitudes has consistently found Southerners to be more likely to acknowledge racial intolerance and to express less enthusiasm for equal opportunity initiatives (Fossett and Kiecolt 1989; Giles 1977; Taylor, Sheatsley & Greeley 1978). It might simply be that racial attitudes affect employment inequality according to conventional expectations and that, all else equal, regional culture fosters greater employment inequality in the South. This might then be counteracted by the inequality-reducing effects of other regional differences. Unfortunately, the lack of area-specific measures of white racial attitudes for a large sample of metropolitan areas prevents us from including direct measures of racial attitudes in the main multivariate analysis reported below. We do however provide a supplementary analysis that attempts to address this issue.

Data and Measures

The measurement of employment inequality and the nature of the metropolitan sample were described earlier. We now briefly describe the operationalization of independent variables included in our regression analysis of employment inequality. A measure of racial education differences was developed from metropolitan statistical area (MSA)-specific census tabulations. For each MSA, educational attainment was coded on a seven-point scale as follows: 0-4 years = 0, 5-7 years = 1, 8 years = 2, 9-11 years = 3, 12 years = 4, 13-15 years = 5, and 16+ years = 6. (This coding follows census educational categories, which emphasize the importance of key credentialing years).[15] Following Fossett and South (1983), we measured inequality by calculating the difference in mean education between whites and blacks and then expressing it as a percentage of the mean education of the group with the higher level of education (in our sample, this was whites in every city).[16]

Labor demand was measured by population change in percentage between 1970 and 1980.[17] Population change is a close proxy for labor force growth, an indicator of long-term labor demand. The cross-sectional correlation between labor force and population exceeds 0.99. This strong correlation justifies using

population growth as a surrogate measure for labor force growth — a considerable practical advantage since it is easier to obtain measurement for population over time based on constant-area MSA definitions.

Prevalence of large firms was measured by the percentage of firms employing 20 or more workers. This is one of the few firm-level indicators that is reported for all industries in the metropolitan areas as opposed to only firms engaged in manufacturing. County-level data were taken from County Business Patterns for 1980 and were aggregated to the metropolitan level. Results for this variable are similar when other cut points were used (e.g., the percentage of firms employing 50 or more workers) and when we used measures of median firm size developed from the firm size distributions.

Union membership data for 1980 come from the U.S. Statistical Abstract's compilation of data from the U.S. Bureau of Labor Statistics. It is measured at the state level and assigned to metropolitan areas (MSAs spanning state borders are assigned an average score). While this is by no means an ideal measurement strategy, alternative MSA-level measures of unionization (e.g., CPS-based estimates from Kokkelenberg and Sockell 1985) are also problematic since they are available for only a small number of areas and are based on area-specific samples of individuals, which are often small.[18] The assignment of state-level unionization scores to metropolitan areas obviously introduces measurement error. However, more reliable estimates are not available and the measurement error introduced should be largely random rather than systematic.[19] Thus, it will tend to attenuate the estimated effect of unionization.

The favorability of the occupational composition to black employment was estimated by using census occupation tabulations to calculate the "expected" black representation in the labor force given the metropolitan area's occupational mix. This was obtained by computing the product of the proportion of the metropolitan area labor force in a given occupation and the percentage representation of blacks in that occupation at the national level and summing these products over all occupations.[20]

The potential impact of race differences in age structure on race differentials in employment was estimated by using census data to calculate the expected employment-to-population ratios for each race in each metropolitan area based on age distributions for each group and national race-age-labor force status tabulations.[21] The expected employment-to-population ratio for whites in the metropolitan area was then expressed as a ratio of the similar figure for blacks and then subjected to a log transformation to obtain an expected relative log-odds of employment based on black-white differences in age composition.

Findings

Table 2 reports the OLS regression of the relative log-odds of employment (i.e., the natural logarithm of the ratio of the white odds of employment to the black odds of employment) on the predictors in the model. The model explains most intermetropolitan variance in racial employment inequality (approximately 70% as indicated by the adjusted R^2). More important, the independent variables all exhibit statistically significant effects in the predicted directions. That is to

say, relative white advantage in employment is greater in areas where race differences in age structure favor white employment, education differences between whites and blacks favor whites, population growth is slow, large firms are prevalent, union membership is high, and representation of occupations favorable to black employment is low. In general then, the predictions about area variation in racial employment inequality developed from human capital theory, queuing theory, and dual labor market theory and from expectations about the impact of occupation mix and race differences in age structure are supported. Two variables, unionization and firm size, appear to be especially important — at least as their respective standardized regression coefficients (betas) of 0.511 and 0.463 seem to indicate. Population growth also stands out as being an important predictor with a standardized regression coefficient of -0.316.

Explaining Regional Differences

We now turn to consider the central question of this article. That is, "What explains the pattern of regional variation in racial employment inequality?" Table 3 presents data on regional differences in the characteristics of metropolitan areas that bear directly on this question.

Considerable insight into regional differences in employment inequality can be gained by examining the data in Table 3 in conjunction with the regression results presented in Table 2. A large regional difference in the means for a given metropolitan area characteristic combined with a sizable coefficient for that characteristic in the regression would suggest a potentially important contribution to regional differences in racial employment inequality. With this in mind, three regional differences in metropolitan area characteristics — namely, population growth, unionization, and firm size — are most important in generating the regional differences in employment inequality observed in Table 1.

Population growth had a negative effect on employment inequality in the regression analysis and metropolitan economies in the Sun Belt were growing rapidly. Average population increases over the decade were approximately 23% in the South and 29% in the West. In contrast, metropolitan economies in the Frost Belt were stagnant, as indicated by average population increases near zero in the Northeast and the Midwest. These patterns contributed to higher employment inequality in the Northeast and Midwest and lower inequality in the South and West. Regional differences in unionization had similar consequences. Unionization had a positive effect on employment inequality, and unionization was decidedly lower in the South and higher in the Northeast and Midwest; the West tracked in between but closer to the Frost Belt. The most striking comparison, of course, is the South versus the Frost Belt, for union membership in the South was lower by about half (15% compared to 30%). Regional differences in firm size were less dramatic. However, the positive effect of firm size on employment inequality contributed to higher inequality in the Midwest, and lower inequality in the West, Northeast, and South.

TABLE 2: OLS Regression of White-Black Male Employment Gap on Demographic, Educational, and Labor Market-Related Factors[a]

	Regression Unstandardized	Coefficient Standardized	t
Age structure differences	1.57	0.166	2.64
Racial education gap	0.0147	0.149	5.77
Population growth	-0.00486	-0.316	-4.03
Traditional black occupational mix	-0.131	-0.158	-2.60
Percent of firms ≥20 employees	0.0810	0.463	7.55
Unionization	0.0150	0.511	6.23

R^2 = .71
Adjusted R^2 = .71
N = 99

[a] U.S. SMSAs 1980. All coefficients are significant at the .05 level in a two-tailed test. The constant is not presented but is used in predicted values in subsequent tables.

The other regional differences in metropolitan area characteristics either had no major consequences for employment inequality or had consequences that ran counter to the general pattern of regional differences in employment inequality. Regional differences in both age structure and occupational mix were modest in size and generally did not have important or systematic impacts on regional differences in inequality. Regional variation in racial education differences are considerable, but the consequences for inequality did not contribute to regional differences in inequality in a simple way. The average metropolitan area in the South exhibited a white advantage in education over blacks of almost 24 percentage points compared to only 15 or 16 in the Northeast and Midwest. The West was lower still, with an average below 11%. Thus, racial education differences contributed to higher employment inequality in the South, lower inequality in the Northeast and Midwest, and considerably lower inequality in the West.

Table 4 presents a more systematic quantification of the contribution of regional differences in the characteristics of metropolitan areas to regional variation in racial employment inequality. The first row of the table presents the regional means of the dependent variable (i.e., the relative log-odds of employment) as deviations from the mean for the national sample. This shows that metropolitan areas in the Northeast and especially the Midwest are well above the national average on racial inequality in employment. By comparison, metropolitan areas in the South and West are well below the national average.

The next six rows of the table (under "Percent of Difference Due to") list the contribution of each region's departure from the national average for each characteristic to its departure from the national average for employment

TABLE 3: Means of Predictor Variables by Region[a]

	Northeast	Midwest	South	West	Total
Age structure differences	-.0879	-.0712	-.0762	-.1012	-.0811
Racial education gap	16.26	15.39	23.93	10.59	18.70
Population growth	-1.176	2.883	23.087	29.039	15.292
Traditional black occupational mix	9.567	9.762	9.596	9.340	9.587
Percent of firms ≥20 employees	13.96	15.63	14.37	13.91	14.49
Unionization	31.24	30.54	15.11	26.37	23.02
N	18	21	45	15	99

$R^2 = .71$
Adjusted $R^2 = .71$

inequality. They were calculated in the following manner. First, the regional mean (in deviation form) of each independent variable was multiplied by the unstandardized regression coefficient for that predictor reported in Table 2. This represents the region's expected deviation from the national average on inequality associated with the region's deviation from the national average on the independent variable in question. This figure was then divided by the observed regional mean of the dependent variable (again in deviation form).[22]

The result is the contribution of the regional differences in metropolitan area characteristics to regional differences in employment inequality expressed as a percentage of the observed regional deviation on employment inequality. The conversion of the contribution of each regional deviation to a percentage interpretation greatly facilitates comparisons of the importance of variables. For example, the greater population growth in southern metropolitan areas contributes 32.88% of the deviation of metropolitan areas in the South of -0.115 from the national average on the relative log-odds of employment. This is not inconsiderable, but it is dwarfed by the impact of the deviation of Southern metropolitan areas on unionization, which generates a deviation on inequality that represents fully 103.29% of the observed southern deviation from the national average on inequality.

The bottom row of the table provides the sum of the separate contributions. If regional differences were entirely captured by these independent variables, the separate contributions of each variable would sum to exactly 100% of the observed regional deviation. Sums below 100% indicate that the model underpredicts the regional deviation on employment inequality. Sums above 100% indicate that the model overpredicts the regional deviation on employment inequality.

Inspection of Table 4 reveals that most of the Northeast's high inequality can be attributed to its high unionization and low population growth. At

TABLE 4: Regression-based Decomposition of Regional Differences in White-Black Employment Gap

	Positive Deviations		Negative Deviations	
	Northeast	Midwest	South	West
Log-relative odds of white-black employment rate: regional deviation from national mean	0.128	0.230	-0.115	-0.131
Percent of difference due to				
Age structure differences	-8.37	6.72	-6.60	24.19
Racial education gap	-28.05	-21.15	-66.73	91.34
Population growth	62.34	26.20	32.88	51.15
Traditional black occupational mix	2.04	-9.96	1.02	-24.77
Percent firms ≥20 employees	-33.44	40.04	8.77	36.07
Unionization	96.30	49.16	103.29	-38.54
Sum of separate contributions	90.82	91.00	72.63	139.44
N	18	21	45	15

contributions of 96.3% and 62.34%, respectively, these two components dwarf the other metropolitan area characteristics considered in the analysis. These contributions to higher average inequality are offset somewhat by the presence of small firms and a certain equality in educational opportunity.

Higher average inequality in the Midwest stems primarily from above-average unionization and firm size. Below-average population growth makes a lesser contribution to inequality. The only major factor reducing inequality is the region's relative educational equality.

As noted previously, the relatively low employment inequality for metropolitan areas in the South can be traced to its below-average unionization and above-average growth, unionization being the far more important factor. The only other factor of consequence is the much higher educational inequality in the South, which greatly increases employment inequality and more than offsets the ameliorating effects of growth.

The pattern in the West does not resemble that observed for any other region. Very slight racial inequality in education contributes more than 90% to the region's observed deviation on employment inequality. Sizable contributions to lower average inequality can also be traced to high average population growth and small average firm size. Higher unionization and an occupational mix unfavorable to black employment contribute to higher employment inequality, but these contributions are modest in size and hardly counterbalance the large effects of low education differences and high growth in the West.

Overall, the model in Table 2 does a good job of reproducing the regional variations in average employment inequality. The predictions are very close for

the Northeast and Midwest and close for the West and South. In the Northeast and Midwest the predicted regional deviations on inequality sum to about 91% of the observed deviations for these regions. These figures are very close to the ideal of exactly 100%. The model underestimates the extent of Southern employment inequality, and overestimates the amount of Western employment inequality. However, predicting 73% of the Southern differential and 139% of the Western differential can be viewed as acceptable performance. Figure 1 shows these differences graphically. It confirms that, in general, the model is helpful in understanding how regional differences in racial employment inequality are generated.

A formal assessment of our model's ability to explain regional differences in racial employment inequality is presented in Table 5. This table presents five regression analyses. The first is our "core model." The second is a model that includes dummies for region only, omitting the Midwest. The third is a model that adds region dummies to our core model. The fourth and fifth models repeat the analysis of region using a two-category Sun Belt-Frost Belt scheme, omitting the Frost Belt in panel E.

Our analytic model explains most of the effects of region in the four-region analysis and all the effects of region in the Frost Belt-Sun Belt analysis. As seen in panel B, region has large and important effects when considered in isolation. However, the explanatory power of the model increases dramatically and regional effects are sharply attenuated when our substantive predictors are added to the model. In panel C, the adjusted R^2 increases to 0.72 (compared to 0.30 in panel B), the deviations associated with the Northeast and the West are near zero, and the deviation for the South is only slightly greater than half its value in panel B. By comparison, the addition of the regional dummies to the core model leads to only a small increase in ability to explain variation in racial employment inequality (the adjusted R^2 increases from 0.71 to 0.72 between panels A and C).

Figure 1 illustrates this goodness of fit graphically. The observed regional deviations in the white-black relative log-odds of employment are charted against those deviations that would be expected from the model in Table 2. The observed and expected graphs are very similar. Clearly, only a small portion of the variation in racial employment inequality can be uniquely attributed to regional effects not captured by our substantive variables.

An Alternative Specification with Racial Attitudes

One objection that can be raised against the analysis presented in Table 2 is that it does not take account of discrimination based on race prejudice. As we argued above, it is not immediately obvious that inclusion of such a variable is important to accounting for regional variation in racial employment inequality since the literature on racial attitudes shows higher levels of racial tolerance in the Northeast and Midwest where observed employment inequality is high (Fossett & Kiecolt 1989; Giles 1977; Middleton 1976; Taylor, Sheatsley & Greeley 1978; Wilson 1986). Nevertheless, we do not dismiss the possibility that normative climates reflected in racial attitudes may have an independent effect

TABLE 5: OLS Regression of White-Black Male Employment Gap on Demographic, Educational, and Labor Market-Related Factors and Regional Dummies[a]

	Regression Coefficient Unstandard.	Standard.	t
	Panel A: Core Only		
Age structure differences	1.57	0.166	2.64
Racial education gap	0.0147	0.149	5.77
Population growth	-0.0049	-0.316	-4.03
Traditional black occupational mix	-0.131	-0.158	-2.60
Percent of firms ≥20 employees	0.0810	0.463	7.55
Unionization	0.0150	0.511	6.23
R^2 = .71			
Adjusted R^2 = .71			
N = 99			
	Panel B: Four Regions, Midwest Omitted		
Northeast			
without core	-0.102†	-.148	-1.43
with core	0.001†	.001	0.013
South			
without core	-0.345	-.648	-5.87
with core	-0.183	-.343	-3.07
West			
without core	-0.361	-.488	-4.80
with core	-0.050†	.067	0.80
	Without Core	With Core	
R^2	.32	.75	
Adjusted R^2	.30	.72	
N = 99			
	Panel C: Four Regions plus Core		
Age structure differences	1.85	0.196	3.14
Racial education gap	0.0184	0.524	6.84
Population growth	-0.0053	-0.343	-3.91
Traditional black occupational mix	-0.096*	-0.116	-1.93
Percent of firms ≥20 employees	0.0797	0.456	7.18
Unionization	0.0077	0.026	2.55
R^2 = .75			
Adjusted R^2 = .72			
N = 99			

TABLE 5: OLS Regression of White-Black Male Employment Gap on Demographic, Educational, and Labor Market-Related Factors and Regional Dummies[a] (Continued)

	Regression Coefficient		
	Unstandard.	Standard.	*t*
	Panel D: Two Regions plus Core		
Age structure differences	1.66	0.175	2.78
Racial education gap	0.0142	0.406	5.57
Population growth	-0.0041	-0.266	-3.12
Traditional black occupational mix	-0.133	-0.160	-2.65
Percent of firms ≥20 employees	0.0793	0.454	7.39
Unionization	0.0130	0.440	4.64
R^2 = .71			
Adjusted R^2 = .69			
N = 99			
	Panel E: Two Regions, Frost Belt Omitted		
Without core	-0.302	-.557	-6.60
With core	0.071†	0.131	-1.45
R^2 = .31			
Adjusted R^2 = .30			
N = 99			

[a] U.S. SMSAs, 1980. All other coefficients are significant at the .05 level in a two-tailed test. The constant is not presented.

† Not significant

* $p < .10$

on hiring decisions that might, if adequately captured, alter the results just reviewed.

We investigated this issue by obtaining average scores on white racial tolerance at the community level generated using data from the 1976 and 1977 General Social Surveys (Davis & Smith 1978).[23] These years were selected because they contained relevant questions on racial attitudes and because community-level information was available (Stinchcombe & Taylor 1980) which permitted identification of 55 metropolitan areas, 45 of which were included in our sample. Note that because data were available only for 45 cities, it was impossible to include these materials in the primary 99-city analysis. Racial tolerance was measured for individuals based on a four-item index similar to those used in previous analyses of racial attitudes (e.g., Fossett & Kiecolt 1989; Taylor, Sheatsley & Greeley 1978; Wilson 1986), but limited to items that

TABLE 6: OLS Regression of White-Black Male Employment Gap on Demographic, Educational, and Labor Market-Related Factors and White Racial Tolerance[a]

	Regression Coefficient Unstandard.	Standard.	t
Age structure differences	0.694†	0.065	0.69
Racial education gap	0.0163	0.452	4.86
Population growth	-0.0049	-0.339	-2.72
Traditional black occupational mix	-0.279	-0.158	-2.60
Percent of firms ≥20 employees	0.0735	0.390	4.59
Unionization	0.0193	0.636	5.06
White racial tolerance	1.71†	0.081	0.78

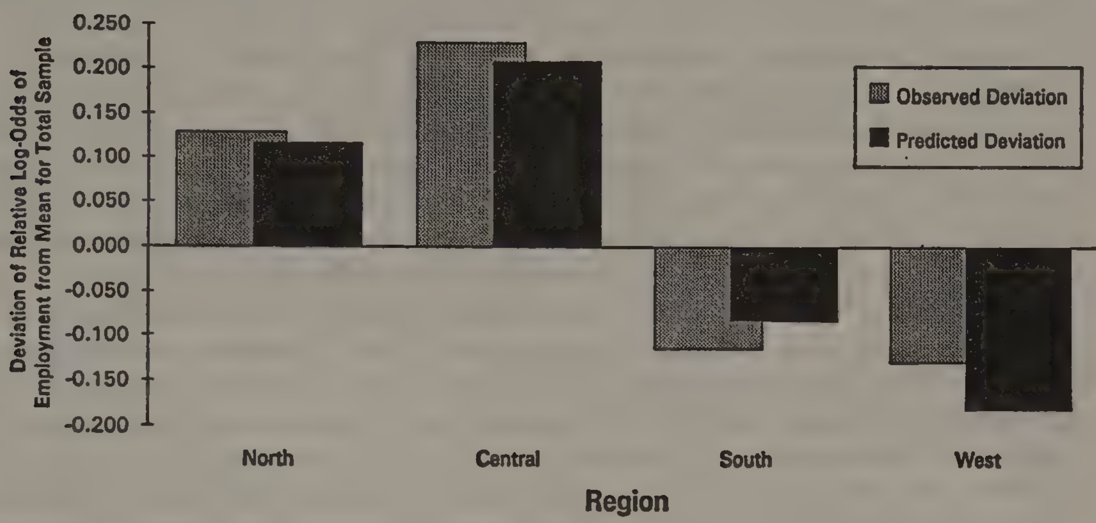
R^2 = .79
Adjusted R^2 = .75
N = 45

[a] U.S. SMSAs 1980
† Not significant. All other coefficients are significant at the .05 level in a two-tailed test. The constant is not presented.

measure adherence to the principle of racial integration (Schuman, Bobo & Steeh 1985). Respondents were asked whether whites have a right to racial residential segregation, whether blacks and whites should go to the same schools or to different schools, whether there should be laws against marriages between blacks and whites, and whether blacks should "push" themselves where they are not wanted. Individual responses to these four items were coded so higher scores indicated greater tolerance or acceptance of integration and were summed to create an index with scores ranging from 4 to 12 with higher values indicating greater support for racial integration.[24]

Area-specific means of the individual index scores were computed for each metropolitan area included in the GSS sample. The average level of white racial tolerance varied meaningfully and significantly across metropolitan areas ranging from a low of 6.07 for Birmingham, Alabama, to a high of 9.66 for San Francisco. Unfortunately, despite combining two years of the GSS, the area-specific samples of individuals were not large (e.g., the median sample size over the 45 metropolitan areas in our sample was 26) and the racial tolerance means were less than wholly reliable (the median standard error of the area-specific racial tolerance means was 0.41). However, this measure of racial tolerance varied across areas in ways that appeared sensible based on a working knowledge of race relations in the U.S. (e.g., lower scores in the South and higher scores in areas with high levels of education and income) Thus, while imperfect, it is useful and represents the best available direct measurement of white racial attitudes in different metropolitan areas known to us.

FIGURE 1: Observed versus Predicted Regional Differences in the White-Black Relative Log-Odds of Employment



We added the racial tolerance means as an independent variable in our data set and reestimated the regression equation presented in Table 2 using the smaller sample of 45 cases for which the racial tolerance means could be computed. The results of this analysis are presented in Table 6. Two important findings warrant discussion. The first is that racial tolerance has no significant impact on employment inequality. The effect of white attitudes is both substantively small and statistically insignificant. The second finding is that most of the effects of the other predictors in the model are unaffected when racial attitudes are included as a control. With the exception of age structure differences, every variable maintains its sign, approximate size, and significance. Age structure becomes insignificant, but this variable had already been found to contribute little to our substantive understanding of regional differences in employment inequality. Variables that were previously identified as having important consequences for interregional variation in racial employment inequality — economic growth, unionization, firm size, and educational differentials — continue to have similar effects under this new specification.[25]

Of course, we must be careful not to dismiss the role of racial attitudes on the basis of this relatively slender evidence. The measures of racial attitudes, while useful, leave much to be desired both in terms of coverage of areas and reliability for a given area. Nevertheless, evidence for a central role of racial attitudes in explaining regional differences in employment discrimination or altering our understanding of the effects of other factors cannot be adduced from this analysis.

Conclusion

In this article we have shown that metropolitan areas in the Northeast and the Midwest are characterized by higher levels of racial inequality in employment than metropolitan areas in the South and West. We also have shown that a number of theoretically relevant metropolitan characteristics predict area variation in racial employment inequality well and that regional differences in employment inequality are closely tied to regional differences on these metropolitan characteristics. Specifically, our analysis suggests that greater employment inequality observed in the Midwest and Northeast can be attributed to the lower economic growth, larger firm size, and higher union membership found in these regions compared to the South and West.

The finding that employment inequality is greater in the Northeast and Midwest is interesting because it contrasts with previous research focusing on other important indicators of racial inequality. Analyses of racial inequality in education, occupation, and income consistently document greater inequality in the South.[26] We can only speculate why interarea variation in racial employment inequality follows a different logic. Education differences are likely to be the function of racial dynamics in the funding and operation of school systems, factors that may not be germane to hiring dynamics. Occupation and income inequality may be lower in the North than the South because unions and the internal labor markets associated with large firms may provide protection against wage discrimination for employed workers. Unions and large firms may thus reduce racial inequality among workers who obtain jobs.[27] However, these protections may perversely intensify racial inequalities in employment by increasing the exclusion of minorities from employment in the first place.

In any event, at least one other important indicator of racial inequality — namely residential segregation — varies across regions according to the pattern observed here for employment inequality. Massey and Denton (1987, 1993) found that levels of white-black residential segregation were lower in the South and West and higher in the Northeast and Midwest (labeled North Central in their tables). This result also been reported by Farley and Frey (1994). Another parallel between our analysis and these analyses is that they both present evidence that residential segregation tends to decline in areas that experience strong economic growth (which is accompanied by growth in population and housing stock). Specifically, Massey and Denton found that employment growth had a positive effect on residential contact between whites and blacks and a negative effect on residential dissimilarity between blacks and whites. Similarly, Farley and Frey report that recent growth in housing stock has a negative effect on residential dissimilarity. The strong parallel between our findings and their findings in this regard provides further confirmation that a simple cultural explanation (focusing on prejudice) for racial inequality is likely to be an insufficient explanation of interregional variation in racial inequality.

A more complete explanation of differences in the regional patterns for various dimensions of racial inequality is needed. Future analyses should consider multiple dependent variables to determine whether regional differences in racial inequality in education, occupation, and income are simply subject to different determinants or whether their variation across areas can be understood

within a general framework that also accounts for area variation in employment inequality and residential segregation.

We hope that this article will help stimulate further consideration of the regional and area differences in racial opportunity. Regional and other interarea variation in racial inequalities has been somewhat neglected in the literature in recent decades. Theoretical perspectives such as the human capital and status attainment traditions have emphasized the labor supply side of stratification processes. They have focused attention on the characteristics individuals bring to labor markets and thus on the role of group differences in education and related resources in shaping inequality. Insights from these perspectives are important and they contribute to our understanding of inequality. However, inordinate focus on the effects of individual-level characteristics can sometimes obscure the fact that important macro patterns are there to be observed in the analysis of racial inequality.

These patterns suggest contextual effects operating in labor markets, regions, the state, and other macro units. Understanding the effects of employer practices, state policies, union strength, aggregate labor demand, and other factors shaping the nature of labor demand and its expression ultimately requires examination of macro empirical patterns. Comparative studies of inequality in regions and metropolitan areas provide one analytic tool that can be used to pursue these issues. Studies using other macro units (e.g., firms and organizations) also can be fruitful, although data are often less readily available. The results presented here provide additional evidence suggesting that demand-side factors are particularly important in the study of racial and ethnic inequality where individual outcomes are dramatically shaped by larger economic and political forces ranging from discrimination to economic growth.

Notes

1. These are the metropolitan areas for which detailed tabulations of employment status by age and sex are published for blacks.

2. Thus, in the typical city in the Midwest and Northeast whites' odds of employment are about 3.1 times as great as blacks' odds of employment, whereas in the typical city in the South and West it is only about 2.3 times as great.

3. It is important to remove the institutional and military populations from the calculation of employment-population ratios. Institutional populations (e.g., inmates of prisons) and military populations can have dramatic effects on employment-population ratios, yet their employment status does not reflect local labor market opportunities. Moreover, these populations do not necessarily originate in the local area and thus no simple relationship between local opportunities and the relative size of institutional and military populations can be assumed.

4. That is, $(W_e/W_n)/(B_e/B_n)$ where W_e is the number employed for whites, W_n is the number not employed for whites, B_e is the number employed for blacks, and B_n is the number not employed for blacks (all counts are for civilian, noninstitutional males 25-59 years old).

5. In contrast, the major alternative measure of employment inequality — the white-black difference in employment rates — is not well suited for regression analysis because standard OLS assumptions of linearity in relationships and normal distribution of the error terms are violated for these dependent variables (Hanushek & Jackson 1977). As is well known, these violations are nontrivial because they can lead to bias in estimates of causal effects and invalid tests of statistical significance (Aldrich & Nelson 1984). These potential problems are avoided

APPENDIX A: Log Relative Odds of White to Black Male Employment Rates (Age 25-59)

Northeast (N = 18)		*Midwest (N = 21)*	
Philadelphia, Pa.	1.307	Grand Rapids, Mich.	1.452
Harrisburg, Pa.	1.295	Chicago, Ill.	1.415
Newark, N.J.	1.286	Milwaukee, Wisc.	1.350
Buffalo, N.Y.	1.263	Detroit, Mich.	1.341
Paterson, N.J.	1.254	Fort Wayne, Ind.	1.330
Trenton, N.J.	1.200	Indianapolis, Ind.	1.305
Pittsburgh, Pa.	1.188	Cleveland, Ohio	1.268
Rochester, N.Y.	1.159	Saint Louis, Mo.	1.261
Springfield, Mass.	1.051	Wichita, Kan.	1.215
New York, N.Y.	1.023	Youngstown, Ohio	1.184
Boston, Mass.	1.006	Omaha, Neb.	1.133
Syracuse, N.Y.	1.004	Cincinnati, Ohio	1.125
Bridgeport, Conn.	0.984	Akron, Ohio	1.111
Long Branch, N.J.	0.980	Toledo, Ohio	1.100
Albany, N.Y.	0.927	Minneapolis, Minn.	1.087
U.S. median	*0.927*	Dayton, Ohio	1.044
New Brunswick, N.J.	0.722	Gary, Ind.	1.036
Jersey City, N.J.	0.633	Kansas City, Mo.	1.016
Nassau, N.Y.	0.602	*U.S. median*	*0.927*
		Ann Arbor, Mich.	0.902
		Flint, Mich.	0.852
		Lansing, Mich.	0.639
Median	1.048	Median	1.133

[a] SMSAs containing conurbations, such as Dallas-Ft. Worth, have been shortened for space to the name of the first city, for example, Dallas.

when employment differentials are modeled in terms of the log-odds and relative log-odds (Aldrich & Nelson 1984; Hanushek & Jackson 1977).

6. We are careful not to overemphasize this point, as the racial difference in employment rates can be a useful measure of employment inequality in some situations. Fortunately, it is highly correlated with the relative log-odds ratio in the our data ($r = 0.90$) and it describes a similar pattern of interregional variation in racial employment inequality.

7. The census sample records are drawn at the household level and the sampling rate may vary across small areas within the metropolitan area. The fraction of total persons represented in the sample data for a metropolitan area thus varies but rarely falls outside the range of 15 to 20%. (For further details see 1980 Census of Population, Volume 1 — Characteristics of the Population, Chapter D — Detailed Characteristics, Appendix D.)

APPENDIX A: Log Relative Odds of White to Black Male Employment Rates, Age 25-59 (U.S. SMSAs 1980)[a] (Continued)

South ($N = 45$)		*West* ($N = 15$)	
Baltimore, Md.	1.201	Fresno, Calif.	1.328
Memphis, Tenn.	1.183	San Francisco, Calif.	1.142
Montgomery, Ala.	1.114	Portland, Ore.	1.063
Jackson, Miss.	1.100	*U.S. median*	*0.927*
Atlanta, Ga.	1.066	Los Angeles, Calif.	0.923
Beaumont, Miss.	1.037	Seattle, Wash.	0.844
Washington, D.C.	1.031	Phoenix, Ariz.	0.761
Richmond, Va.	1.007	Sacramento, Calif.	0.759
Birmingham, Ala.	1.007	Denver, Colo.	0.744
Macon, Ga.	0.995	San Jose, Calif.	0.681
Lexington, Ky.	0.987	Vallejo, Calif.	0.668
Baton Rouge, La.	0.978	San Diego, Calif.	0.663
Houston, Tex.	0.945	Tacoma, Wash.	0.649
New Orleans, La.	0.942	Riverside, Calif.	0.603
U.S. median	*0.927*	Las Vegas, Nev.	0.575
Tulsa, Okla.	0.924	Anaheim, Calif.	0.449
Dallas, Tex.	0.920		
Mobile, Ala.	0.899		
Jacksonville, Fla.	0.892		
Shreveport, La.	0.889		
Louisville, Ky.	0.877		
Columbia, Mo.	0.870		
Charlotte, N.C.	0.870		
Chattanooga, Tenn.	0.841		
Charleston, W. Va.	0.837		
Wilmington, N.C.	0.836		
Newport News, Va.	0.801		
Norfolk, Va.	0.781		
Nashville, Tenn.	0.773		
Oklahoma City, Okla.	0.767		
Little Rock, Ark.	0.765		
Greenville, S.C.	0.743		
Huntsville, Ala.	0.729		
Greensboro, N.C.	0.717		
Austin, Tex.	0.696		
San Antonio, Tex.	0.625		
W. Palm Beach, Fla.	0.590		
Knoxville, Tenn.	0.589		
Orlando, Fla.	0.552		
Ft. Lauderdale, Fla.	0.501		
Miami, Fla.	0.481		
Pensacola, Fla.	0.477		
Augusta, Ga.	0.470		
Lakeland, Fla.	0.373		
Tampa, Fla.	0.349		
Daytona Beach, Fla.	0.226		
Median	0.841	Median	0.744

[a] SMSAs containing conurbations, such as Dallas-Ft. Worth, have been shortened for space to the name of the first city, for example, Dallas.

8. Four cities — Hartford, Connecticut; New Haven, Connecticut; Raleigh-Durham, North Carolina; and Columbus, Ohio — are excluded because of missing data on independent variables used in the regression analyses.

9. Non-Hispanic whites are not tabulated directly, since the census category "whites" includes the Spanish origin population. We obtained data for non-Hispanic whites by subtracting nonwhite groups and Spanish origin from the total as appropriate. One minor problem with this procedure is that blacks of Spanish origin may be subtracted twice. However, we can find no evidence that this ever distorts the employment patterns for non-Hispanic whites in any significant way.

10. We use the median and the interdecile range to describe central tendency and dispersion for the relative odds ratios because distributions of odds ratios tend to be right skewed, rendering means and standard deviations inappropriate for descriptive statistics.

11. Many variables that are important determinants of inequality (e.g., mode of economic organization, nature of the political order, etc.) are simply not relevant for the task at hand because they are effectively constants across regions and metropolitan areas of the U.S. and therefore cannot account for regional differences in inequality.

12. Of course, this is an oversimplification since the investment theory that undergirds human capital theory implies that group differences in investments in education will be shaped by group differences in returns to education in the labor market.

13. Whether race is in fact a reliable proxy for such unmeasured productivity characteristics is an empirical question. What is clear, however, is that under conditions of slack labor demand, statistical discriminators will not necessarily suffer from ranking potential employees on the basis of ethnicity and may potentially gain. For example, unless race is inversely associated with unmeasured productivity characteristics, employers who use race as a tie-breaker to rank applicants with otherwise similar qualifications will at worst obtain employees who are no less qualified than those hired by nondiscriminating employers.

14. We currently are undertaking studies of mismatch effects based on detailed data on job and residential locations. However, our analysis will be limited to about a dozen cities due the limited availability of the relevant data and the cost of performing each within-city analysis.

15. An alternative based on years of schooling completed (estimated from category midpoints) proved less effective.

16. Conceptually, this is an attractive formulation because, unlike the more commonly used ratio of mean education, it is "symmetrical" in depicting relative advantage to whites and blacks (Fossett & South 1983). This can be important in samples where advantage sometimes favors blacks and sometimes favors whites. In our sample, however, educational inequality always favors whites and as a result the two measures are highly correlated ($r > 0.99$). Thus, analyses using the ratio of mean education to measure educational inequality yield identical substantive results.

17. We also examined unemployment rates. This had an effect similar to population growth, but the effect was less important and less robust than the effect of population growth.

18. We also considered state-level unionization estimates developed by Kokkelenberg and Sockell (1985). They produced virtually identical results. This is not surprising, since their correlation with the state-level estimates we used exceeded 0.97.

19. Systematic bias would be introduced by this measurement strategy only if MSA-specific deviations from state unionization scores were correlated with other variables in the analysis or with other potentially important but omitted variables.

20. That is, the expected black representation in the labor force based on the occupational mix is given by $\Sigma O_i P_i$ where O_i is the proportion of the labor force in the metropolitan area in occupation i and P_i is the percentage black in the occupation at the national level.

21. For example, the expected employment-to-population ratio for blacks in a given metropolitan area was given by $\Sigma A_i P_i$ where A_i is the proportion of blacks in the metropolitan area in age group i and P_i is the percentage employed for blacks in age group i at the national level.

22. For example, average population growth for metropolitan areas in the South was 23.087. This deviates from the national average of 15.292 by 7.795 points. The unstandardized regression coefficient for population growth reported in Table 2 was -0.00486. Multiplying 7.795

by -0.00486 yields -0.03788, which is the contribution of population growth to the South's deviation from the national average for the relative log-odds of employment inequality. This represents 32.88% of the observed regional deviation from the national average for inequality of -0.11522. That is 32.88 = 100.0 * -0.03788 / -0.11522. (Note the calculations reported here are based on numbers that are more precise than those reported in the tables. Performing the same calculations using the reported numbers will introduce rounding error.)

23. We are indebted to K. Jill Kiecolt who generously provided us with these data.

24. Individual-level factor analysis of the items in the racial tolerance scale yielded a single-factor solution. Cronbach's alpha is 0.66.

25. In fact, the metric effects (i.e., the unstandardized regression coefficients) of unionization, population growth, educational inequality, and occupational mix increase under this specification.

26. The pattern for employment inequality in our study cannot be attributed to an unusual sample. We examined data on racial inequality in aggregate attainments in education, occupation, and family income for metropolitan areas in our sample. As reported in previous studies, the South was characterized by higher levels of inequality for these indicators.

27. Relative to this point, it is worth noting that studies of occupational inequality focus on the employed population or the experienced civilian labor (without regard for employment status). Similarly, studies of income inequality often focus on comparisons involving persons who are employed or even employed full-time and year-round. Cotton (1989) has shown that analyses focusing on per capita income may yield decidedly different results.

References

Aldrich, John H., and Forrest D. Nelson. 1984. *Linear Probability, Logit, and Probit Models*. Sage Publications.

Arrow, Kenneth. 1972. "Some Mathematical Models of Discrimination in the Labor Market." Pp. 187-204 in *Racial Discrimination in Economic Life*, edited by Anthony Pascal. D.C. Heath.

Ashenfelter, Orley. 1971. "Discrimination and Trade Unions." Pp. 83-118 in *Discrimination in Labor Markets*, edited by Orley Ashenfelter and Albert Rees. Princeton University Press.

Averitt, Robert T. 1968. *The Dual Economy: Dynamics of American Industry Structure*. Norton.

Becker, Gary S. 1957. *The Economics of Discrimination*. University of Chicago Press.

Braddock, Jomills Henry, and James M. McPartland. 1987. "How Minorities Continue to Be Excluded from Equal Opportunities: Research on Labor Market and Institutional Barriers." *Journal of Social Issues* 43:5-39.

Burman, George. 1973. "The Economics of Discrimination: Impact of Public Policy." Ph.D. thesis, University of Chicago.

Cohn, Samuel R. 1985. *The Process of Occupational Sex-Typing: Feminization of Clerical Labor in Great Britain*. Temple University Press.

Cotton, Jeremiah. 1989. "Opening the Gap: Decline in Black Economic Indicators in the 1980s." *Social Science Quarterly* 70:803-19.

Darden, Joe T. 1983. "Racial Differences in Unemployment: A Spatial Perspective." *Review of Black Political Economy* 12:93-106.

Davis, James A., and Tom W. Smith. 1978. *Cumulative Codebook for the 1972-1978 General Social Surveys*. University of Chicago, National Opinion Research Center.

Farley, John E. 1987. "Disproportionate Black and Hispanic Unemployment in U.S. Metropolitan Areas: The Roles of Racial Inequality, Segregation and Discrimination in Male Joblessness." *American Journal of Economics and Sociology* 46:129-50.

Farley, Reynolds. 1984. *Blacks and Whites: Narrowing the Gap*. Harvard University Press.

Farley, Reynolds, and Walter R. Allen. 1987. *The Color Line and the Quality of Life in America.* Russell Sage.

Farley, Reynolds, and William H. Frey. 1994. "Changes in the Segregation of Whites from Blacks during the 1980s: Small Steps toward a More Integrated Society." *American Sociological Review* 59:23-45.

Fogel, Walter A. 1970. *The Negro in the Meat Industry.* Wharton.

Foner, Philip S. 1974. *Organized Labor and the Black Worker, 1619-1973.* Praeger.

Fossett, Mark A., and K. Jill Kiecolt. 1989. "The Relative Size of Minority Populations and White Racial Attitudes." *Social Science Quarterly* 70:820-35.

Fossett, Mark A., Omer R. Galle, and William R. Kelly. 1986. "Racial Occupational Inequality, 1940-1980: National and Regional Trends." *American Sociological Review* 51:421-29.

Fossett, Mark, and Scott J. South. 1983. "The Measurement of Inter-Group Inequality in Income: A Conceptual Review." *Social Forces* 61:855-71.

Freeman, Richard B. 1973. "Changes in the Labor Market for Black Americans, 1948-1972." Brookings Papers on Economic Activity. Vol. 1. Brookings Institution.

Freeman, Richard B., and James L. Medoff. 1984. *What Do Unions Do?* Basic Books.

Galster, George C. 1987. "Residential Segregation and Interracial Economic Disparities: A Simultaneous Equations Approach." *Journal of Urban Economics* 21:22-44.

Giles, Michael W. 1977. "Percent Black and Racial Hostility: An Old Assumption Reexamined." *Social Science Quarterly* 58:412-17.

Hanushek, Eric A., and John E. Jackson. 1977. *Statistical Methods for Social Scientists.* Academic Press.

Heckman, James J., and Brook S. Payner. 1989. "Determining the Impact of Federal Antidiscrimination Policy on the Economic Status of Blacks: A Study of South Carolina." *American Economic Review* 79:138-77.

Heckman, James J., and Kenneth J. Wolpin. 1976. "Does the Contract Compliance Program Work: An Analysis of Chicago Data?" *Industrial and Labor Relations Review* 29:544-64.

Heer, David M. 1959. "The Sentiment of White Supremacy: An Ecological Study." *American Journal of Sociology* 64:592-98.

Hill, Herbert. 1984. "Race and Ethnicity in Organized Labor: Historical Sources of Resistance to Affirmative Action." *Journal of Intergroup Relations* 12:5-49.

Hill, Richard Child. 1974. "Unionization and Racial Income Inequality in the Metropolis." *American Sociological Review* 39:507-22.

Hodge, Robert W. 1973. "Toward a Theory of Race Differences in Employment." *Social Forces* 52:16-31.

Hodson, Randy D. 1978. "Labor in the Monopoly, Competitive, and State Sectors of Production." *Politics and Society* 8:429-80.

Ihlandfeldt, Keith, and David L. Sjöquist. 1989. "Impact of Job Decentralization on the Economic Welfare of Central City Blacks." *Journal of Urban Economics* 26:110-30.

Jacobson, Julius. 1968. *The Negro and the American Labor Movement.* Anchor.

Jacoby, Sanford M. 1985. *Employing Bureaucracy: Managers, Unions and the Transformation of Work in American Industry 1900-1945.* Columbia University Press.

Jaynes, Gerald David, and Robin M. Williams Jr. (eds.). 1989. *A Common Destiny: Blacks and American Society.* National Academy Press.

Jencks, Christopher. 1991. "Is the American Underclass Growing?" Pp. 28-100 in *The Urban Underclass,* edited by Christopher Jencks and Paul Peterson. Brookings Institution.

Kain, John F. 1968. "Housing Segregation, Negro Employment and Metropolitan Decentralization." *Quarterly Journal of Economics* 82:175-97.

____. 1974. "Reply to Stanley Masters' Comment on 'Housing Segregation, Negro Employment and Metropolitan Decentralization.'" *Quarterly Journal of Economics* 88:513-19.

Kalleberg, Arne L., Michael Wallace, and Robert P. Althauser. 1981. "Economic Segmentation, Worker Power and Job Inequality." *American Journal of Sociology* 87:651-83.

Kasarda, John D. 1990. "Urban Industrial Transition and the Underclass." *Annals of the American Academy of Political and Social Science* 501:26-47.

Kimeldorf, Howard S. 1988. *Reds or Rackets: The Making of Radical and Conservative Unions on the Waterfront*. University of California Press.

Kirshenman, Joleen, and Kathryn M. Neckerman. 1991. "'We'd Love to Hire Them but . . .': The Meaning of Race for Employers." Pp. 203-32 in *The Urban Underclass*, edited by Christopher Jencks and Paul E. Peterson. Brookings Institution.

Kokkelenberg, Edward C., and Donna R. Sockell. 1985. "Union Membership in the United States, 1973-1981." *Industrial and Labor Relations Review* 38:497-543.

Lieberson, Stanley. 1980. *A Piece of the Pie: Black and White Immigrants since 1980*. University of California Press.

Marshall, Ray. 1965. *The Negro and Organized Labor*. Wiley.

Massey, Douglas S., and Nancy A. Denton. 1987. "Trends in the Residential Segregation of Blacks, Hispanics and Asians 1970-80." *American Sociological Review* 52:802-25.

____. 1993. *American Apartheid*. Harvard University Press.

Middleton, Russ. 1976. "Regional Differences in Prejudice." *American Sociological Review* 41:94-117.

Myrdal, Gunnar. 1944. *An American Dilemma: Negro Problem and Modern Democracy*. Harper.

Northrup, Herbert R. 1944. *Organized Labor and the Negro*. Harper.

Phelps, Edmund. 1972. "The Statistical Theory of Racism and Sexism." *American Economic Review* 62:659-61.

Reich, Michael. 1981. *Racial Inequality: A Political-Economic Analysis*. Princeton University Press.

Reskin, Barbara F., and Patricia A. Roos. 1990. *Job Queues, Gender Queues*. Temple University Press.

Roof, Wade Clark, Thomas L. van Valey, and Daphne Spain. 1976. "Residential Segregation in Southern Cities: 1970." *Social Forces* 55:59-71.

Schuman, Howard, Lawrence R. Bobo, and Charlotte Steeh. 1985. *Racial Attitudes in America: Trends and Interpretations*. Harvard University Press.

Shulman, Steven. 1987. "Discrimination, Human Capital and Black-White Unemployment: Evidence from Cities." *Journal of Human Resources* 22:361-76.

____. 1991. "Why Is the Black Unemployment Rate Always Twice as High as the White Unemployment Rate?" Pp. 5-38 in *New Approaches to Economic and Social Analyses of Discrimination*, edited by Richard Cornwall and Phanindra V. Wunnava. Praeger.

Slichter, Sumner. 1960. *The Impact of Collective Bargaining on Management*. Brookings Institution.

Sørensen, Annemette, Karl E. Taeuber, and Leslie J. Hollingsworth Jr. 1975. "Indexes of Racial Residential Segregation for 109 Cities in the United States, 1940 to 1970." *Sociological Focus* 8:128-30.

Stinchcombe, Arthur L., and D. Garth Taylor. 1980. Contextual Analysis Project [Machine-readable data file]. University of Chicago, National Opinion Research Center.

Taylor, D. Garth, Paul B. Sheatsley, and Andrew M. Greeley. 1978. "Attitudes toward Racial Integration." *Scientific American* 238:42-39.

Thurow, Lester C. 1975. *Generating Inequality: Mechanisms of Distribution in the U.S. Economy*. Basic Books.

Vroman, Wayne. 1975. "Changes in the Labor Market Position of Black Men since 1964." Pp. 294-301 in *Proceedings of the Twenty-seventh Annual Winter Meeting, 1974*, edited by James Stern and Barbara Dennis. Industrial Relations Research Association Series. Industrial Relations Research Association.

Wallace, Michael, and Arne Kalleberg. 1981. "The Economic Organization of Firms and Labor Market Consequences: Towards a Specification of Dual Labor Market Theory." Pp. 77-117 in *Sociological Perspectives on Labor Markets*, edited by Ivar Berg. Academic Press.

Welch, Finis R. 1967. "Labor Market Discrimination: An Interpretation of Income Differences in the Rural South." *Journal of Political Economy* 75:225-40.

Wilcox, Jerry, and W. Clark Roof. 1978. "Percent Black and Black-White Status Inequality: Southern versus Non-Southern Patterns." *Social Science Quarterly* 59:421-34.

Williams, Donald R. 1987. *Labor Force Participation of Black and White Youth*. UMI.

Wilson, Thomas C. 1986. "Interregional Migration and Racial Attitudes." *Social Forces* 65:177-86.

Composition of the Workplace and Psychological Well-Being: The Effects of Tokenism on America's Black Elite*

PAMELA BRABOY JACKSON, *Duke University*
PEGGY A. THOITS, *Vanderbilt University*
HOWARD F. TAYLOR, *Princeton University*

Abstract

Kanter's theory of proportional representation suggests that tokens should experience more work stress and psychological symptoms than nontokens. We examine the effects of proportional representation by race and by gender on work stress and symptoms. Data come from structured personal interviews with a disproportionate stratified sample of elite black leaders in the U.S. (N = 167). Consistent with expectations, analyses showed that numerical rarity by race and by gender significantly increased symptoms of depression and anxiety, respectively. Numerical rarity by race significantly increases "token stress" (e.g., loss of black identity, multiple demands of being black, sense of isolation, having to show greater competence) and a high degree of gender tokenism increases role overload. Some, but not all, of the total impact of proportional representation is mediated through work stressors since these stressors are themselves directly associated with higher psychological symptoms.

Numerical representation is an influential structural characteristic within most formal work organizations. Research suggests that minority-group size affects attitudes, achievement, and the frequency and quantity of interpersonal contact between majority and minority group members (Izraeli 1983; Konrad, Winter & Gutek 1992; South et al. 1982; Toren & Kraus 1987). Most of this research draws upon Kanter's (1977) theory of proportional representation, which argues that individuals who occupy token positions in their work settings experience three sources of stress: performance pressures, boundary heightening, and role entrapment.

* *The research reported here was supported by a grant from the Cornerhouse Fund, New York, NY.* *Direct all correspondence to Pamela Braboy Jackson, Duke University, Department of Sociology, Box 90088, Durham, NC 27708-0088.*

 Social Forces, December 1995, 74(2):543-557

There are three limitations to this research that we wish to address in this study. First, the hypothesis that proportional representation determines these distinct sources of work stress has not been systematically examined. Second, studies generally focus their attention on gender tokens. As a result, the implications of Kanter's hypothesis for other underrepresented groups have been neglected. And third, studies concentrate on tokens within single occupations without providing a general view of the effects of minority-group size on token experiences across work settings (see Wharton & Baron 1987 for an exception).

The purpose of this study is to examine the relationship between proportional representation and work-related difficulties. Although Kanter (1977) focuses on the consequences of particular work experiences for tokens' job performance, we argue that these stressors should affect tokens' psychological well-being as well. Relationships are examined using a sample of men and women drawn from the black leadership elite in America. Our extension of Kanter's theory should lead to a more comprehensive understanding of the ways in which the structure of relationships within the work setting can affect individual well-being.

Kanter's Theory of Proportional Representation

Discussion of the effects of differences in proportional representation within collectivities is not unique to Kanter. Hughes (1944, 1946, 1958) described the potential perpetuation of status conflict based on stereotyped expectations for behavior, highlighting the particular dilemmas faced by women and blacks who enter male and white occupations, respectively (also see Jackson 1962). Taylor and her colleagues (Taylor & Fiske 1975; Taylor et al. 1978) have focused on the evaluation of numerical minorities in white male-dominated groups. Other observational studies on numerical distinctiveness have been conducted (see Segal 1962; Wolman & Frank 1975; Yoder & Sinnett 1985) and currently the literature on the impact of proportional representation on individual group members is growing (e.g., Abrams, Thomas & Hogg 1990; Alexander & Thoits 1985; Macke 1981; Ruble & Higgins 1978; Spangler, Gordon & Pipkin 1978; Wharton & Baron 1987; Yoder 1994). The area of inquiry is not new, but Kanter provides a useful theoretical framework for it. Her emphasis on the sex composition of the group is important, but it is her elaboration of the stress produced by differences in status characteristics and tokens' reactions to these problems that guides our inquiry here.

According to Kanter (1977), work groups can be characterized as uniform, skewed, tilted, or balanced in proportional representation. Uniform groups are homogeneous; all members hold the same master statuses (e.g., white males). The ratio of majority to minority group members in such groups is 100:0. Stress due to differences in observable master status characteristics is absent in these work settings. In skewed groups, majority group members far outnumber minority group members. Kanter suggests ratios of majority to minority group members ranging from 99:1 through 85:15. She refers to majority group members as "dominants" and minority group members as "tokens" in skewed

groups. Those in numerical majority are assumed to control the group and its culture; tokens have very little power in these situations. Tokens are "often treated as representatives of their category, as symbols rather than individuals" (Kanter 1977:208). Tilted groups represent less extreme distributions. Kanter suggests ratios from 84:16 to 65:35, where dominants are still the majority and tokens still the minority. However, due to greater numbers, minority group members are one another's potential allies, can form coalitions, and can affect the culture of the group. As a result, Kanter hypothesizes that minority group members do not experience as much stress as those in skewed groups. Finally, in balanced groups, proportions range from 64:36 to 50:50. Culture and interaction, she argues, become balanced between majority and minority subgroups. In this situation, individual outcomes depend upon structural and personal factors related to the group member rather than group composition per se.[1]

According to the theory of proportional representation, tokens in skewed groups are the most vulnerable to lower levels of emotional well-being from three sources of stress. One discussed by Kanter (1977) is *performance pressure*. Because their "differentness" is highly visible, tokens feel that they are always under scrutiny. Further, because they are symbolic representatives of their "type," tokens experience added pressure to perform well, since this may determine future opportunities for other individuals in their social category. Tokens perceive that their performance ratings are based at least in part on their master status characteristics and not their actual achievements on the job. Thus, tokens may have to put in extra effort to have their achievements noticed. One response to performance pressure is to overachieve; another is to avoid calling attention to oneself or one's accomplishments (and thus potentially to be perceived as less competent or underachieving).

The second source of stress, *boundary heightening*, results from majority group members' tendencies to exaggerate their own commonalities as well as their differences from tokens. Tokens are repeatedly reminded of their difference through jokes, interruptions, exclusion from informal activities, and various "loyalty tests." As responses, tokens can either remain socially isolated or present themselves as exceptions to their category in order to become accepted insiders.

The third source of stress, *role entrapment*, involves typecasting by dominants. Tokens' behaviors are assimilated into existing cultural stereotypes of women (or blacks). Dominants' stereotyped assumptions and mistaken attributions force tokens to play limited and caricatured roles in the organization (e.g., for women, "mother" and "seductress"). Those who resist stereotyped roles are "trapped in a more militant stance than they might otherwise take" (Kanter 1977:236). Either way, "the personal consequence for tokens, of course, is a certain degree of self-distortion" (984).

We assume (as does Kanter) that performance pressures, boundary heightening, and role entrapment are experienced as stressors, i.e., as ongoing difficulties on the job that require behavioral readjustments. Further, tokens' adaptive responses — working harder or "laying low" and thereby suffering lack of recognition, accepting social isolation or turning against one's own (and thus, to some extent, oneself), playing stereotyped roles or combating stereo-

types — are likely to be stressors as well. Consequently, we expect token members in skewed groups to report higher levels of these stressors and of psychological symptoms than those in tilted, balanced, and other groups in which their "type" are the majority.

Kanter's work focused on the situations facing token women in sales management positions and male/female interaction patterns within a large industrial organization. We assume that the dynamics of interaction between "tokens" and "dominants" will be similar across work contexts, at least for leading professionals, government administrators, and chief executives (on whom we focus our analysis here). Tokens are identified by ascribed characteristics (i.e., gender, race, ethnicity). Attached to these characteristics are sets of secondary and informal assumptions about the culture, competence, and behavior of the status occupant. Thus, any interaction will be altered because attention is directed to the master status characteristic or the stereotypes informally attached to the status rather than to the individual's ability to perform a specific task (Hughes 1944; Kanter 1977). In this article, we focus on work stressors associated both with racial and gender tokenism.

Data and Methods

The data for this study come from a research project directed by the third author at Princeton University in 1983. The project was to investigate the network structure of America's black leadership. The primary goals of the study were to collect systematic background information on black elites, to examine the structure of the network links among these leaders, and to describe variations in the political and policy attitudes of the group. Structured personal interviews also elicited information on the gender and race composition of respondents' workplaces, occupational stress, coping responses, and symptoms of psychological distress. There were 167 respondents in the sample.

SAMPLE

To obtain a sampling frame of America's national black leaders, several population lists were compiled. First, institutional sector lists were developed (see Taylor 1992 for a detailed annotation of sources for each institutional sector). Sectors included major executives in both predominantly white and predominantly black businesses, then-current black mayors of major cities (and some former mayors), then-current black members of the U.S. Congress, black federal judges, high-ranking military officers, heads of predominantly black voluntary and civil-rights organizations, black heads of major philanthropic foundations, major media personalities (including columnists, correspondents, editors of major magazines, editors of black newspapers in large cities, entertainers, and athletes), high-ranking labor union officers, major civil service officials, and high-ranking Democratic, Republican, and independent party officers. Other listings included presidents of predominantly black colleges and universities, nationally prominent ministers, scholars and educators, writers and literary figures (intellectuals), and wealthy blacks.

Second, blacks in *Who's Who in America* (WWA; 1982) were enumerated. Because WWA does not indicate individuals' race or ethnicity, *Who's Who among Black Americans* (WWBA; 1980) was used as a cross-check for race. WWBA is by far the longest source list of prominent blacks in the U.S. (17,000 names were included in the 1980-81 edition). There were 960 names common to the two sources.[2] These names were supplemented by lists of national "influentials" published by *Ebony* magazine periodically since 1963 and yearly since 1971.[3] Those names not already on the WWA/WWBA list were added to it.

The WWA/WWBA list was compared to the institutional sector lists; all names on the WWA/WWBA list that did not appear in the institutional sector lists were added to the appropriate institutional listing. Generally, disproportionate (equal frequency) stratified sampling by institutional sector was employed (15-20 persons were drawn from each sector list), with random sampling within sectors. However, *all* black members of the U.S. Congress in 1984 and black heads of major philanthropic foundations were purposively included, given the small number of such individuals in each category. A list of 260 potential respondents resulted from these sampling procedures.

The response rate in the study was 70.5% ($N = 167$), excluding those who, between the time of sampling and contact for interview, had died, moved from the U.S., or were unreachable for other reasons ($N = 23$). This response rate is comparable to rates obtained in other major studies of elites, which range from 40% to 80%, with 60% to 70% being modal (Taylor 1992). The majority of respondents were interviewed by black interviewers.[4]

Table 1 reports the background characteristics of respondents. The majority are middle-aged, male, married, highly educated (95% are college graduates, 63% have advanced degrees), report median family earnings of $82,000 per year, and are in a wide range of prestigious occupations.[5]

Measures

PSYCHOLOGICAL DISTRESS

Abbreviated symptom scales from the Johns Hopkins Symptom Checklist were used to assess depression and anxiety. For brevity, items were selected if they had high factor loadings on depression and anxiety (Derogatis et al. 1971) and also matched symptom items included in the National Survey of Black Americans (Jackson 1979). Exploratory factor analyses showed that these items formed two factors in this sample. Depression was indicated by the sum of four items: "In the past week, how often did you feel down or depressed, feel hopeless about the future, lack enthusiasm for doing anything, feel lonely?" Cronbach's alpha for this scale was .75. Anxiety consisted of three items: feel tense or keyed up, feel nervous, have your heart pound or race when you were not physically active ($\alpha = .77$). Responses for these items ranged from 1 = never to 5 = very often.

TABLE 1: Descriptive Characteristics of Black Elites

Variable		N
Age		158
Mean age	54.6	
Median age	53.0	
Gender		167
Percent male	76.6	
Marital status		161
Percent married	77.0	
Education		152
Mean years of education	18.1	
Percent some college education	95.4	
Income		127
Mean household income 1985	$107k	
Median household income 1985	$82k	
Occupation		167
Civil servant	15.0	
Major executive	12.6	
Congressperson	10.2	
Labor leader	10.2	
Military leader	8.4	
Foundation head	7.8	
Leader of interest group	7.2	
Intellectual	7.2	
Media personality	6.6	
College president	4.2	
Mayor	3.6	
Political party official	1.8	
Civil rights leader	1.2	
Financial institution	1.2	
Entertainer	1.2	
Religious leader	1.2	

PROPORTIONAL REPRESENTATION

Two questions assessed occupational group composition for race and for gender. "On a typical work day, do you work with people all of whom are black (female), most of whom are black (female), about half of whom are black (female), most of whom are white (male), all of whom are white (male) except you, or what?" Possible responses were coded: 5 = all are black (female);

4 = most are black (female); 3 = about half are black (female); 2 = most are white (male); 1 = all are white (male) except respondent. Follow-up questions asked, "About what percent would you say are black (female)?" These questions yielded both *ordinal* and *continuous percentage* indicators of black and female proportional representation. As a third measure, continuous percentages were collapsed using Kanter's suggested cut points, distinguishing skewed (1-15%), tilted (16-35%), and balanced (36-50%) groups (which we refer to below as "Kanter's cut points"). For all three indicators of proportional representation, lower values indicated situations in which whites (or males) are numerically dominant, higher values indicated black (or female) numerical dominance in the workplace. For regression analyses, for women respondents only, the percentage of female representation in the workplace was subtracted from 100 to yield the percentage of male representation. Thus, in regressions, the continuous gender percentage measure becomes the percentage of persons in the workplace of the *opposite gender* from the respondent (for both male and female respondents).

A majority of respondents reported working with mostly white and mostly male coworkers. The ordinal measures and Kanter's cut points placed 55-58% of the respondents in white-dominated groups (i.e., in skewed and tilted groups). About 46% of the respondents were in male-dominated workplaces according to the ordinal measures and 50% according to Kanter's cut points. Correlations among the three measures of proportional representation by race were quite high.[6] Given the consistency among these measures, we use the continuous percentage measure below in regressions.

A weakness of these proportional representation questions is their failure to distinguish explicitly between coworkers and clients. Respondents might be including the percentage of black (or female) clients with whom they come into contact each day. However, Kanter (1977) quite clearly states that the interpersonal dynamics that she discusses (performance scrutiny, boundary heightening, role entrapment) occur in contacts with clients as well as colleagues. Consequently, the lack of specificity in these questions should not be a serious drawback and may reflect the realities of relative numerical contacts in the workplace.

WORK STRESSORS

Respondents were asked if their work involved any of 16 pressures. These items included problems identified by Kanter: scrutiny, performance pressures, nonacceptance, and social isolation.[7] Several items tapped another potential stressor not discussed by Kanter, feelings of overload. Tokens presumably are more likely to be asked by their institutions and more generally by their public to serve as symbolic representatives of their "type" on additional committees, on panels, in policy sessions, and so forth. Consequently, tokens should experience multiple demands due to being black, have more demands made on their time in general, and have difficulties juggling their work and private lives.

Respondents indicated whether they experienced each pressure on the job (0 = no, 1 = yes), and if so, how much this pressure bothered them (1 = not at all, 2 = somewhat, 3 = very much). The frequency distributions of how much the pressure bothered them were bimodal, with 1 and 3 most frequent. Consequent-

ly, we retained dichotomous indicators for each work stress item (coded 0 = problem is not experienced, 1 = problem is experienced).

Exploratory factor analyses suggested five types of work stress. They included *scrutiny*: scrutiny by colleagues and public scrutiny; *nonacceptance*: experiencing discrimination, value conflicts, nonacceptance by whites, and nonacceptance by blacks; *token stress*: loss of black identity, multiple demands of being black, having to demonstrate more competence than peers, and a sense of isolation; *interpersonal conflict*: hostility from others, political infighting, and conflicting demands from various groups; and *role overload*: too many time demands, juggling private and work life, and too many responsibilities. Items composing each factor were summed as indices of each type of work stress. Because occupations differ greatly in structure (e.g., some jobs involve public scrutiny but not scrutiny by colleagues), we view these work stress measures not as scales but as indices, i.e., as sums of experienced work problems, grouped by type (Heise 1974).

It should also be noted that these questions tap individuals' subjective perceptions of occupational stressors rather than objective conditions of work. Thus, there may be some confounding of perceived work stress with individuals' psychological symptoms (e.g., highly depressed respondents might perceive more sources of occupational stress). We assume here that these successful, highly influential respondents are able to report fairly accurately on the stressors they encounter in their professional lives, but the possibility of reverse causality cannot be dismissed. See Appendix for correlations between work stressors and psychological symptoms.

Background variables include age (measured in years), gender (0 = male; 1 = female), marital status (0 = unmarried; 1 = married), education (measured in years), income (measured in thousands of dollars), and occupation (four dummy variables representing businesspersons, academics, entertainers, and "other" leaders; with politicians as the omitted comparison category).[8]

Results

TOKENISM AND PSYCHOLOGICAL DISTRESS

We expected black leaders who are outnumbered by whites in their work situations to exhibit higher levels of distress than those in balanced situations and situations in which blacks outnumber whites. Similarly, we expected men and women who are outnumbered in their workplaces by the opposite sex to have higher symptom scores than those who are surrounded by equal or greater numbers of the same sex. To examine these hypotheses, we regressed depression and anxiety scores on degree of racial and gender tokenism, holding background variables constant. The results from these baseline models show that our expectations were generally substantiated. As shown in panel A of Figure 1, the degree of race tokenism is marginally significantly related to higher depression scores, while high opposite-gender representation in the workplace is significantly associated with greater anxiety. Squared terms for race and gender representation were then added to each equation to check for possible curvilinearities. Curvilinearity was found in the relationship between

black representation and anxiety (the beta for percent black was 4.83, $p < .10$; the beta for percent black squared was -6.28, $p < .05$), indicating lower symptoms of anxiety at the highest levels of black representation (not shown in Figure 1).[9]

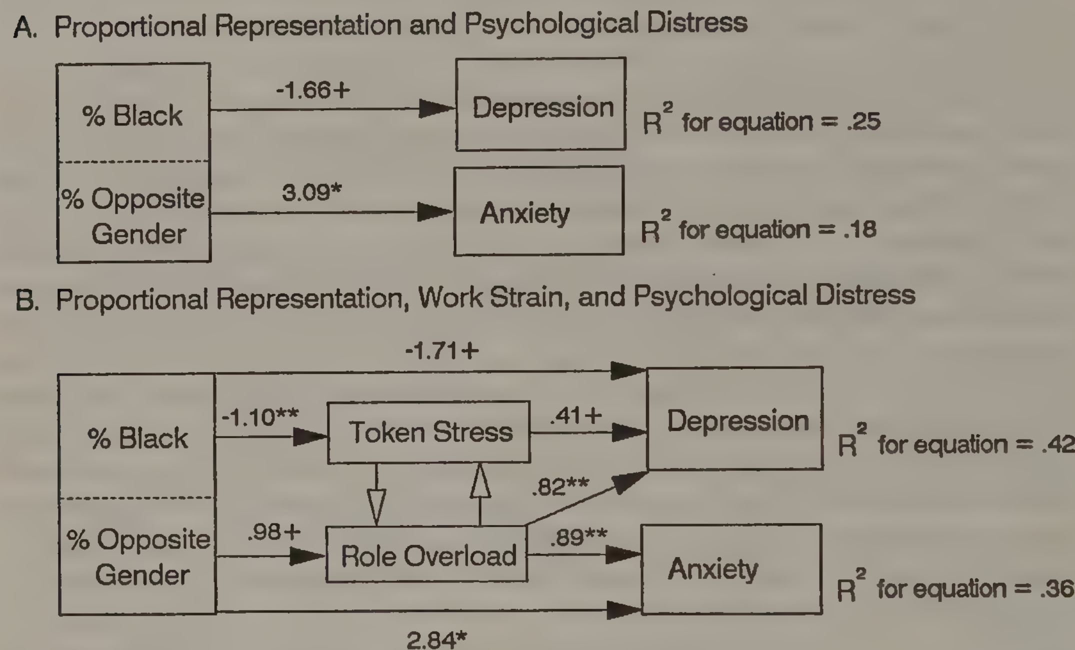
TOKENISM AND WORK STRESS

According to the theory of proportional representation, blacks in white-dominated workplaces, men in female-dominated workplaces, and women in male-dominated workplaces should report more work stressors than others whose colleagues and clients share similar master status characteristics. To test this hypothesis, each work stress index was regressed on the proportional representation measures, the other indices of work stress, and the background control variables. The left half of panel B in Figure 1 shows that race proportionality has a significant effect on token stress. Elites who work in predominantly black work settings report fewer problems with token stress than elites in work settings that are not predominantly black, a finding consistent with Kanter's (1977) hypothesis. However, proportional representation by race is not associated with other kinds of work stresses (scrutiny, nonacceptance, interpersonal conflict, role conflict). Elites in workplaces dominated by members of the opposite gender, on the other hand, report significantly more problems with role overload than their peers. Again, however, representation by gender is unrelated to other types of work stress. Thus, we found only partial support for Kanter's general hypothesis that same-status representation in the workplace affects the experience of work problems. That support is limited to certain status compositions of the workplace and particular kinds of occupational problems.[10]

TOKENISM, WORK STRESSORS, AND PSYCHOLOGICAL DISTRESS

The final step in the analysis examined the relationships among proportional representation, work stressors, and psychological symptoms. Does the experience of specific work stressors explain the higher depression and anxiety scores of elites who are outnumbered by whites and the opposite sex, respectively? Depression and anxiety scores were regressed on the proportional representation measures, the five types of work stress, and the background control variables. The results are presented in the right half of panel B, Figure 1. As shown, the degree of race tokenism continues to be a significant predictor of depression, even when token stress and other work stressors are controlled. Similarly, the direct effect of opposite-gender representation on anxiety remains statistically significant when role overload and other stressors are held constant. Again we see that proportional representation in the workplace has significant impacts on psychological well-being, as initially hypothesized. Moreover, these impacts are both direct and indirect through specific work stressors. Elite blacks who are outnumbered by whites report more token stress and token stress, in turn, increases depression. Elite men and women who are outnumbered by the opposite gender report more role overload and role overload, in turn, increases both depression and anxiety.[11] When squared terms for percent black and percent opposite-gender representation were added to these equations, again the inverse U-shaped relationship between percent black and anxiety appeared (the

FIGURE 1: Regression Results — Symptoms and Work Stressors for Race and Gender Tokens[a]



(N = 124)

[a] Unstandardized coefficients are presented. See text for other variables controlled in each equation.

+ $p < .10$ * $p < .05$ ** $p < .01$

B for percent black was 4.23, p = ns; the *B* for percent black squared was -6.39, $p < .05$), indicating lower anxiety at the highest levels of black representation, net of work stressors (not shown in Figure 1).[12]

Discussion and Conclusions

Kanter's (1977) theory of proportional representation suggests that individuals who hold token positions will experience distinct problems at work compared to peers working in settings that are balanced in status composition or dominated by members of the same gender or race. Kanter's analysis also implies that tokens should experience greater symptoms of psychological distress due to work pressures. This article provided an empirical test of these hypotheses within a sample of black elite leaders.

We found limited support for the argument that a variety of work stressors are associated with proportional representation. Consistent with the theory, the higher the black percentage in the work setting, the fewer problems elites

reported due to the salience of their black identity (specifically, loss of black identity, multiple demands of being black, having to demonstrate more competence than peers, and a sense of isolation). In addition, high opposite-gender representation was associated with experiences of role overload. However, problems due to scrutiny, nonacceptance, and interpersonal conflict did not vary with the racial or gender composition of the workplace. It appears that proportional rarity produces a more restricted range of difficulties than Kanter suggested, at least within this elite sample.

The finding that racial tokenism increases experiences of token stress is consistent with the larger theoretical and empirical literature on gender tokenism (Blau 1977; Kanter 1977; Macke 1981; McGuire et al. 1978), although our gender tokens did not report an excess of these problems (perhaps because a number of the work stress items were worded in ways that referred to racial and not gender identity). Furthermore, both men and women in situations dominated by the opposite sex experienced role overload. This finding contrasts with research that suggests that male tokens are not exposed to the same stressful work conditions as female tokens (Blau 1977; Gutek 1985). We suspect that women and men who are numerically rare in their work settings may be finding themselves in a position where they are often asked to represent their gender "type."[13]

Although the range of difficulties reported by elite tokens was more delimited than expected, these difficulties nevertheless had significant psychological consequences (as did proportional representation itself). Elites experiencing the pressures associated with token status or role overload suffered higher depression or anxiety, respectively. Even when these sources of work stress were controlled, workplace composition directly affected individuals' psychological symptoms. In workplaces dominated by blacks, black elites' symptoms of depression (and anxiety) were lower. In workplaces dominated by the opposite gender, elites' symptoms of anxiety were higher.

In sum, proportional representation by race and by gender did affect elites' psychological well-being both directly and indirectly through work stress. The direct effects of proportional representation on symptoms are important because the phenomenon of tokenism transcends the work setting and influences the dynamics of group interaction more generally. The findings also raise the question of how race and gender tokens cope with the work stressors created or exacerbated by their status. Perhaps effective sources of social support can alleviate the negative impact these particular stressors have on psychological well-being. These issues await future study.

The limitations of this study must, however, be underscored. This was a very select sample of black leaders. Individuals in elite positions represent 1% of the U.S. population and black elites constitute a small fraction of this group. Therefore, the results from this study cannot be generalized beyond such elites. Furthermore, the large number of busy respondents for whom the key variables of interest were skipped is a concern. Finally, the data are cross-sectional, leaving open the possibility of alternative causal orderings of key variables (e.g., token elites may report more sources of work stress if they are depressed or anxious).[14] An important aspect of theory building, nonetheless, involves testing the applicability of the theory to various segments of the population. This

sample of eminent, successful blacks offered a unique opportunity to examine the relevance of Kanter's well-known theory of proportional representation. Additional research will be needed to verify that tokenism, with its accompanying occupational pressures, influences psychological well-being in the general working population. Future research should also explore how elites cope with stressful work conditions.

Notes

1. Kanter offers these ratios as tentative cut points and discusses the concept of "tipping point." Unanswered questions include: How many of a category are enough to change a person's status from token to full group member? When does a group move from skewed to tilted to balanced? We use her suggested ratios as qualitative cut points, but in recognizing that cut points might be set higher or lower, we also examine a continuous measure of proportional representation.

2. Because WWBA includes a large number of locally prominent leaders, it is not an appropriate population list of national influentials. WWA is a key source list because it includes not only government and business leaders but literary figures, academics, entertainers, and sports figures as well (Lieberson & Carter 1979). Our matched list of 960 names is very close to a rough statistical calculation that there are about 1,000 blacks in WWA, which can be derived from estimates in Lieberson and Carter (1979).

3. *Ebony* editors select each year's names. *Ebony* lists tend to overrepresent heads of fraternal organizations and underrepresent literary figures and academics (Henry 1981).

4. Interviews were in person and consisted of both structured and open-ended questions. The average interview lasted two and one-half to three hours (range 1 to 5 hours). Interviews occurred between September 1984 and the spring of 1987, with the majority of interviews completed by the summer of 1985. The response rate varied from sector to sector. Eighty percent of the 21 black members of Congress and 70% of 20 sampled military officers participated. Big-city mayors and federal judges were the most difficult to interview, with response rates of 30% and 45%, respectively, out of 20 in each category. Judges in particular cited the constraints of the law as reasons for refusal.

5. Approximately 30 respondents have missing data on the work stress and psychological symptom scales. Because a number of respondents were extremely pressed for time, interviewers were instructed in such cases to skip specific blocks of interview questions, including psychological distress, work stress, and coping items. Respondents for whom these questions were skipped are quite similar demographically to other respondents in the study. Most are older males (the average age is 55), highly educated (66% report an advanced degree), married (85%), and earn high incomes. Many of these respondents work in Congress ($N = 13$; 43%), suggesting that these politicians were probably the most pressed for time.

6. The ordinal measure of race representation is correlated with the continuous percentage measure at $r = .89$ ($p < .001$) and with the cut points measure at $r = .89$ ($p < .001$). Not surprisingly, the continuous version of the race measure is almost perfectly correlated with Kanter's cut points ($r = .99$, $p < .001$). Similarly high correlations were obtained among the gender indicators. Among men, correlations between the ordinal and interval measures of gender, the ordinal and Kanter measures, and the interval and Kanter measures are .86, .80, and .97, respectively. For women, corresponding correlations are .90, .85, and .98, respectively.

7. Indicators of role entrapment are not in the list; appropriate brief questions tapping this problem were hard to generate.

8. Respondents whose income was missing were assigned the median income category. In addition, there were 14 millionaires in the sample; one reported a net worth of just over 100 million dollars. The income of these respondents was recoded at $50,000 higher than the highest reported income to lessen the impact these outliers would have on our estimates. In

APPENDIX: Correlations of Work Stressors and Psychological Symptoms among Black Elites

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	Mean	S.D.	N
(1) Percent Black	1.00	.12	-.06	.13	-.13	.09	.09	-.09	-.07	-.03	-.07	.35	.30	130
(2) Percent O-Gender		1.00	.02	.08	-.05	-.01	.11	.08	.16	-.05	.05	.44	.23	167
(3) Scrutiny			1.00	.31**	.27**	.33**	.27**	.13	.06	.03	.08	1.45	.73	138
(4) Nonacceptance				1.00	.45**	.51**	.26**	.22*	.17	.15	.21*	1.68	1.46	136
(5) Token stress					1.00	.36**	.27**	.31**	.27**	.28**	.33**	1.74	1.23	133
(6) Interpersonal conflict						1.00	.33*	.10	.15	.09	.13	1.68	1.08	135
(7) Role overload							1.00	.27**	.37**	.31**	.37**	1.88	1.03	138
(8) Depression								1.00	.56**	.58**	.82**	6.69	2.69	137
(9) Anxiety									1.00	.60**	.81**	5.72	2.41	137
(10) Somaticism										1.00	.60**	10.34	3.71	137
(11) Generalized distress											1.00	22.75	7.50	137

* p < .05 ** p < .01

terms of occupational category, politicians included members of Congress, political party officials, civil servants, mayors, civil rights leaders, labor leaders, and leaders of interest groups. Businesspersons included foundation heads, heads of financial institutions, and major executives. Entertainers included actors, athletes, and media personalities. Academics included intellectuals and college presidents. "Other" leaders were in religious and military positions.

9. We also explored the effects of "double tokenism," i.e., being outnumbered both by race *and* by gender in the workplace. Although we have suggestive evidence that double tokens experience higher symptoms of depression and anxiety, the number of these individuals was too small ($N = 17$) to enable meaningful analysis.

10. In each of the regression equations where work stress was the dependent variable, we also examined the impact of proportional representation, *excluding* the other indices of work stress. This approach failed to change the pattern of results reported in the body of the article, therefore, we present the full regression model which controls for other work stressors.

In addition, two interaction terms were added to the equations to assess whether the impact of group representation on work stress (and psychological well-being) differs by gender (i.e., gender*percent black; gender*percent opposite-gender). Only one interaction term was significant. Women who are in predominantly black workplaces report fewer problems with acceptance than other elites ($\beta = -2.01$, $p < .10$).

11. Results from all regression models are available upon request.

12. To further examine the nonlinearities in the effects of proportional representation on anxiety, we regressed anxiety on four dummy variables that indicated the intervals of proportional representation suggested by Kanter (see text), using the balanced group as the comparison category. According to these results, it would appear that the "tilted to black" (i.e., 65:35) category represents the point at which increases in representation begin to reduce anxiety. We would like to thank an anonymous reviewer for recommending this exploration.

13. As noted by one reviewer, this interpretation raises the possibility that the relationship between gender representation in the workplace and psychological well-being may be a gender effect. However, there was no significant relationship between gender and proportional representation. In fact, contrary to expectations, over half of the women work in settings that are either balanced in gender proportions or are predominantly female.

14. We did examine, however, the latter two models holding constant the depression and anxiety scores. The patterns reported in the body of the paper remained significant even under these conditions.

References

Abrams, Dominic, Joanne Thomas, and Michael A. Hogg. 1990. "Numerical Distinctiveness, Social Identity and Gender Salience." *British Journal of Social Psychology* 29:87-92.

Alexander, Victoria D., and Peggy A. Thoits. 1985. "Token Achievement: An Examination of Proportional Representation and Performance Outcomes." *Social Forces* 64:333-40.

Blau, Peter M. 1977. *Inequality and Heterogeneity*. Free Press.

Derogatis, Leonard R., Ronald S. Lipman, Lino Covi, and Karl Rickles. 1971. "Neurotic Symptom Dimensions as Perceived by Psychiatrists and Patients of Various Social Classes." *Archives of General Psychiatry* 24:454-64.

Gutek, Barbara A. 1985. *Sex and the Workplace*. Jossey-Bass.

Heise, David. 1974. "Some Issues in Sociological Measurement." Pp. 1-16 in *Sociological Methodology*, edited by Herbert L. Costner. Jossey-Bass.

Henry, C.P. 1981. "Ebony Elite: America's Most Influential Blacks." *Phylon* 42:120-32.

Hughes, Everett C. 1944. "Dilemmas and Contradictions of Status." *American Journal of Sociology* 50:353-59.

____. 1946. "Race Relations in Industry." Pp. 107-22 in *Industry and Society*, edited by William Foote Whyte. McGraw-Hill.

____. 1958. *Men and Their Work*. Free Press.

Izraeli, Dafna N. 1983. "Sex Effects or Structural Effects? An Empirical Test of Kanter's Theory of Proportions." *Social Forces* 62:153-65.

Jackson, Elton. 1962. "Status Inconsistency and Symptoms of Stress." *American Sociological Review* 27:469-80.

Jackson, James. 1979. National Survey of Black Americans. Unpublished interview schedule, Institute for Social Research, Survey Research Center, University of Michigan, Ann Arbor, Michigan, Project 491722.

Kanter, Rosabeth Moss. 1977. *Men and Women of the Corporation*. Basic Books.

Konrad, Alison M., Susan Winter, and Barbara A. Gutek. 1992. "Diversity in Work Group Sex Composition: Implications for Majority and Minority Members." *Research in the Sociology of Organizations* 10:115-40.

Lieberson, Stanley, and Donna K. Carter. 1979. "Making It in America: Differences between Eminent Blacks and White Ethnic Groups." *American Sociological Review* 44:347-66.

Macke, Anne Statham. 1981. "Token Men and Women: A Note on the Salience of Sex and Occupation among Professionals and Semiprofessionals." *Sociology of Work and Occupations* 8:25-38.

McGuire, William J., Claire V. McGuire, Pamela Child, and Terry Fujioka. 1978. "Salience of Ethnicity in the Spontaneous Self-Concept as a Function of One's Ethnic Distinctiveness in the Social Environment." *Journal of Personality and Social Psychology* 36:511-20.

Ruble, Diane N., and E. Tory Higgins. 1978. "Effects of Group Sex Composition on Self-Presentation and Sex-Typing." *Journal of Social Issues* 32:125-32.

Segal, Bernard E. 1962. "Male Nurses: A Case Study in Status Contradiction and Prestige Loss." *Social Forces* 41:31-38.

South, Scott J., Charles M. Bonjean, William T. Markham, and Judy Corder. 1982. "Social Structure and Intergroup Interaction: Men and Women of the Federal Bureaucracy." *American Sociological Review* 47:587-99.

Spangler, Eve, Marsha A. Gordon, and Ronald M. Pipkin. 1978. "Token Women: An Empirical Test of Kanter's Hypothesis." *American Journal of Sociology* 84:160-70.

Taylor, Howard F. 1992. "The Structure of a National Black Leadership Network: Preliminary Findings." Unpublished paper, Department of Sociology, Princeton University.

Taylor, Shelley E., and Susan T. Fiske. 1975. "The Token in the Small Group: Research Findings and Theoretical Implications." Paper presented to the Yale University Symposium on Psychology and Politics.

Taylor, Shelley E., Susan T. Fiske, Nancy L. Etcoff, and Audrey J. Ruderman. 1978. "Categorical and Contextual Bases of Person Memory and Stereotyping." *Journal of Personality and Social Psychology* 36:778-93.

Toren, Nina, and Vered Kraus. 1987. "The Effects of Minority Size on Women's Position in Academia." *Social Forces* 65:1090-1100.

Wharton, Amy S., and James N. Baron. 1987. "So Happy Together? The Impact of Gender Segregation on Men at Work." *American Sociological Review* 52:574-87.

Who's Who in America. 1982. 42d ed. Marquis.

Who's Who among Black Americans. 1980. 3d ed. Who's Who among Black Americans Publishing Co.

Wolman, Carol, and Hal Frank. 1975. "The Solo Woman in a Professional Peer Group." *American Journal of Orthopsychiatry* 45:164-71.

Yoder, Janice D. 1994. "Looking Beyond Numbers: The Effects of Gender Status, Job Prestige, and Occuapational Gender-Typing on Tokenism Processes." *Social Psychology Quarterly* 57:150-59.

Yoder, Janice D., and Laura M. Sinnett. 1985. "Is It All in the Numbers: A Case Study of Tokenism." *Psychology of Women Quarterly* 9:413-18.

Hispanic Educational Attainment: The Effects of Family Background and Nativity*

ROGER A. WOJTKIEWICZ, *Louisiana State University*
KATHARINE M. DONATO, *Louisiana State University*

Abstract

We first examine the degree to which foreign birth explains the lower educational attainment of Mexicans and Puerto Ricans compared to whites. Although foreign birth is a partial explanation of group differences, family structure and parental education are more important explanatory variables. However, when we consider how the effects of nativity vary across Hispanic groups, we find that while U.S.-born Mexicans have higher educational attainment than foreign-born Mexicans, U.S.-born Puerto Ricans are no better off than foreign-born Puerto Ricans. We also find heterogeneity in the educational experiences of U.S.-born Mexicans. Those with foreign-born parents have higher educational attainment than those with U.S.-born parents.

Dramatic differences in educational attainment between Hispanic and non-Hispanic groups have persisted over the last 20 years. By 1991, only 61% of Hispanics 20 to 21 years old had completed high school compared with 81% of non-Hispanic blacks and 90% of whites (U.S. Department of Education 1992). Because education is a strong predictor of personal income, these differences contribute strongly to explanations of economic inequality between Hispanics and non-Hispanics.

The present study investigates two sources of the difference in educational attainment between Hispanics and non-Hispanics: family background and nativity. As we show below, Hispanics have fewer of the family background characteristics that lead to higher educational attainment. In addition, Hispanics are much more likely to be foreign born, a characteristic related to lower educational attainment. Our study examines the degree to which these differences in family background and nativity explain differences in educational

* *We appreciate helpful comments from Isaac Eberstein, Robert Mare, the editor, and two anonymous reviewers. An earlier version of this article was presented at the 1993 annual meeting of the American Sociological Association.* *Direct correspondence to Roger Wojtkiewicz, Department of Sociology, 126 Stubbs Hall, Louisiana State University, Baton Rouge, LA 70803. Internet: sowojt@lsuvm.sncc.lsu.edu.*

 Social Forces, December 1995, 74(2):559-574

attainment. We also investigate how the negative effect of nativity varies across Hispanic groups.

Previous Research

Strong influences on the educational attainment of Hispanics fall into two types: general and specific (Fernandez, Paulsen & Hirano-Nakanishi 1989). General predictors are important not only for Hispanics but for all race and ethnic groups and include family background factors such as socioeconomic status and aspects of family structure, such as absent parent, presence of a stepparent, and number of siblings.

Prior research on educational attainment has shown that parental education is a strong predictor of children's education (Featherman & Hauser 1978; Haveman, Wolfe & Spaulding 1991; Mare 1979). Parents with more education provide a home environment that supports and encourages the education of children, and they have more income available to finance education and related activities.

Other characteristics have been linked to low educational attainment, such as living in mother-only families (Krein & Beller 1988; McLanahan 1985), living in stepparent families (Astone & McLanahan 1991; Sandefur, McLanahan & Wojtkiewicz 1992; Wojtkiewicz 1993), and having many siblings (Blake 1989). Living with a single parent or stepparent means less available family income and parental time for the socialization, supervision, and control of children. Living with more siblings diminishes the amount of parental and financial resources available to each child (Powell & Steelman 1990).

Some predictors identify characteristics that are much more common among Hispanics than other groups: foreign birth, non-English language background, and short duration of U.S. residence. Studies have documented that proficiency in English lowers the likelihood that Hispanics drop out of high school (Bean & Tienda 1987; Fligstein & Fernandez 1985a, 1985b; Rong & Grant 1992; U.S. Department of Education 1992), whereas not living in the U.S. at the age of 14 raises the chance of dropping out (Rumberger 1983). In addition, foreign birth reduces the likelihood for Hispanics to graduate from high school (Bean & Tienda 1987; Fligstein & Fernandez 1985a, 1985b; Velez 1989). These effects arise because they are linked to lower competency in speaking and understanding English, which hampers academic achievement. They may also indicate educational experiences in countries where the quality of schools is lower, and where levels of final attainment are lower, than in the U.S.

Fernandez, Paulsen, and Hirano-Nakanishi (1989) investigated how these background factors contributed to the difference in educational attainment between Hispanic and non-Hispanic groups by estimating separate logistic regression models for each and then calculating probabilities of dropping out of high school. They found that the gap in educational attainment was reduced after controlling for differences between Hispanic and non-Hispanics in both general and specific characteristics.

A positive aspect of nativity found in other research was that U.S.-born children of immigrants had higher levels of educational attainment than

comparable children of U.S.-born parents (Fligstein & Fernandez 1985a, 1985b; Rong & Grant 1992; U.S. Department of Education 1992). In addition, Kao and Tienda (1995) found that while Hispanic children with foreign-born parents did not have higher grades, mathematics scores, or reading scores, they did have higher college aspirations.

Thus, in contrast to the steady increase that assimilation theorists predict for immigrant socioeconomic status over time, the reality is often quite different Immigrant educational attainment is not simply a product of duration of residence in the U.S.(Portes & Rumbaut 1990)[1]; it may reflect a set of motivations unique to immigrant parents (Fernandez, Paulsen & Hirano-Nakanishi 1989; Lieberson 1980). One consequence may be that children of immigrant parents have better educational outcomes than those of U.S.-born parents.

Present Study

Our analysis explores the degree to which family background and nativity explain the gap between Hispanics and non-Hispanics in educational attainment. Previous research has documented the important role of family background, and nativity is a characteristic that distinguishes Hispanics from other disadvantaged groups, such as blacks and Native Americans.

We begin with a bivariate analysis that shows that (1) educational attainment is related to family background and nativity and (2) family background and nativity are related to race and ethnicity. We then use logistic regression to examine the degree to which educational differences between Hispanics and non-Hispanics are explained by nativity and, subsequently, by other family background characteristics. In the analysis we consider four separate Hispanic groups: Mexicans, Puerto Ricans, Cubans, and other Hispanics. Because these groups are dissimilar in country-of-origin experiences, U.S. settlement patterns, and cultural expectations (Bean & Tienda 1987), we expect considerable variation in their educational attainment.

The next part of the analysis explores the degree to which the effect of nativity varies across Hispanic groups. We are especially interested in whether the effect of nativity differs between Mexicans and Puerto Ricans, who are U.S. citizens by birth and free to enter the U.S. Because Mexicans often enter the U.S. without legal documents (Donato 1994; Massey et al. 1987), illegal status may constrain their educational achievement. Thus, we expect a smaller negative effect of nativity for Puerto Ricans than for Mexicans.

In the final part of our analysis, we differentiate between persons of Mexican ancestry, who were U.S. born with U.S.-born parents, and those born in the U.S. with parents born in Mexico. We expect to replicate earlier studies that report that U.S.-born natives with foreign-born parents attain more education than those with U.S.-born parents. However, we are particularly interested in determining if this difference is due to underlying differences in family background.

Data and Methods

DATA

For this analysis, we use data from the National Longitudinal Survey of Youth. The survey is a national probability sample of men and women who were first interviewed in 1979 when they were between 14 and 21 years old and interviewed every year thereafter. To ensure representativeness and adequate sample sizes, the cross-sectional sample was supplemented with subsamples of groups typically with low representations in national surveys: minority groups, economically disadvantaged non-Hispanic whites, and persons in the military.

The survey is a rich source of information about the labor force and educational transitions that young adults make as they move into adulthood. In addition, it has an excellent response rate. Excluding the military subsample, 90.2% of respondents interviewed in 1979 were interviewed again in 1988 (Center for Human Resource Research 1991). For our purposes, we use the cross-sectional sample and the black and Hispanic supplemental subsamples from the 1979 through 1990 waves of the survey.

DEPENDENT VARIABLE

In this analysis, our dependent variable is high school completion, a key educational transition (Mare 1979, 1981). We are interested in explaining differences in high school completion between race and ethnic groups because having a high school degree provides the basic skills necessary for full-time employment; those who do not obtain a high school degree are much more likely to be in marginal jobs or unemployed. We measure high school completion by whether respondents completed 12 years of school by age 25. Those who had received General Equivalency Degree certificates without a high school diploma were classified as not having a high school degree.

INDEPENDENT VARIABLES

Our independent variables include race and ethnicity, nativity, family structure, socioeconomic status, and parents' nativity. We measured race and ethnicity using survey questions that asked respondents to identify their origin or descent. In cases where more than one origin was mentioned, we used the origin to which respondents reported identifying most closely. This process produced eight categories for race and ethnicity. In addition to blacks, Native Americans, Asians, and non-Hispanic whites, we identify four Hispanic groups: Mexicans, Puerto Ricans, Cubans, and other Hispanics. The latter includes persons with a variety of national origins. By Mexican, Puerto Rican, or Cuban, we mean those who are of Mexican, Puerto Rican, Cuban heritage either by birth or ancestry.

To create the nativity variables, we used information from respondents about whether they, or their parents, were born in the U.S. For those born in the U.S. or its territories, we were able to identify those of Puerto Rican origin who were born in Puerto Rico.[2] Because the survey did not ask the specific place of each parent's birth for those born in U.S. territories, however, we were unable to determine the nativity of the parents of Puerto Rican respondents.

We also include two variables that are closely related to nativity. One variable is whether a foreign language was spoken at home during the respondent's childhood. By including this variable in the analysis, we can determine the degree to which nativity captures the effects of weaker English language skills. The second variable is residence at the age of 14. Including this variable allows us to consider the extent to which the effects of nativity on graduation are due to growing up in countries where dropping out of school at young ages occurs more often than in the U.S. Because no precise measure of residence duration is in the survey, residence at the age of 14 also provides a rough control for the length of time respondents have spent in the U.S.

Finally, we also include parental structure and education and number of siblings in our models because these variables strongly affect educational attainment. Parental structure measures with which parents respondents lived at the age of 14. Parental education measures the educational level of either the mother or father, whichever was higher. We also include a variable for sex in our models to control for any sample differences between race and ethnic groups.

Descriptive Analysis

GRADUATION RATES

Our results are consistent with previous findings and show that Mexicans and Puerto Ricans have the lowest graduation rates (Table 1). Slightly more than half the Puerto Ricans in our sample were high school graduates, whereas 60% of Mexicans graduated from high school. Non-Hispanic whites and Asians were much more likely to graduate from high school (with completion rates over 80%), and about three-quarters of the Cubans, other Hispanics, and blacks completed high school.

The foreign born exhibited lower high school graduation rates than natives. Respondents living in homes where a foreign language was spoken had lower high school completion rates than those in homes without a foreign language. In addition, those living in the U.S. at the age of 14 had higher high school graduation rates than those not.

There were strong differences by family background in high school graduation rates. Those living in a single-parent or stepparent family were less likely to graduate than those living with two biological parents, and those with more siblings were less likely to finish than those with fewer siblings. Almost

TABLE 1: Unweighted Frequencies and Percentages of High School Graduates[a]

	Unweighted Frequency	Percent High School Graduates
Race and ethnicity		
Non-Hispanic white	4,288	84.3
Mexican	1,047	60.5
Puerto Rican	265	54.9
Cuban	96	77.7
Other Hispanic	175	76.8
Black	2,614	72.9
Native American	343	72.5
Asian	66	82.7
Nativity		
U.S. born	8,164	81.3
Foreign born	730	68.3
Foreign language		
No	6,920	81.8
Yes	1,974	73.9
Residence at age 14		
In U.S.	8,659	81.1
Not in U.S.	235	57.8
Sex/Gender		
Male	4,333	78.9
Female	4,561	82.5
Parental structure		
Two parents	6,043	85.3
Single parent	1,762	68.3
Stepparent	709	67.3
Other	380	58.0
Siblings		
0-1	1,408	86.7
2-3	3,409	84.2
4 or more	4,077	73.5
Parental education		
< High school	2,914	62.2
High school	4,347	84.3
College	1,306	95.6
Missing	327	48.5
Total	8,894	80.7

[a] Percentages are calculated using sample weights.

all respondents with college-educated parents graduated from high school. In contrast, less than two-thirds of those with high school dropout parents completed high school.

DISTRIBUTIONS

Compared to non-Hispanic whites, Hispanic groups have consistently higher percentages with backgrounds that lead to lower educational attainment (Table 2). For example, Hispanics are much more likely than non-Hispanic whites to be foreign born. About 26% of Mexicans, 35% of Puerto Ricans, and 58% of Cubans were foreign born compared to just 3% of non-Hispanic whites.

On the whole, Hispanics were much more likely than non-Hispanic whites to live in a home where a foreign language was spoken. Nine percent of whites lived in such a situation compared to 92% of Mexicans, 96% of Puerto Ricans, and 97% of Cubans. Hispanics were also more likely than non-Hispanic whites to have lived outside the U.S. as teenagers. Ten percent of Mexicans, 7% of Puerto Ricans, and 7% of Cubans lived elsewhere at the age of 14, but only 1% of whites did so.

Puerto Ricans and Mexicans were more likely to experience disadvantageous family structures than non-Hispanic whites. For example, 37% of Puerto Ricans and 16% of Mexicans lived in single-parent families at the age of 14 compared to 10% of whites. While 31% of whites reported four or more siblings, almost two-thirds of Mexicans and more than half of all Puerto Ricans had large sibships. Moreover, the educational attainment of Puerto Rican and Mexican parents was also considerably less than whites. Only 16% of whites had parents who completed less than high school compared to 65% of Mexicans and 58% of Puerto Ricans.

Multivariate Analysis

ADDITIVE MODELS

Below we provide a formal test of the degree to which the higher proportion of Hispanics with disadvantaged background characteristics contribute to differences in graduation rates between Hispanics and non-Hispanic whites. We begin our examination in Table 3 by estimating three logistic regression models predicting whether respondents completed high school. Model 1 compares the likelihood of graduating from high school for Mexicans, Puerto Ricans, Cubans, other Hispanics, blacks, Native Americans, and Asians to that for whites, without controlling for other predictors. With the exceptions of Cubans and Asians, each of the race and ethnic groups have significantly lower chances of graduating from high school than whites, and Mexicans and Puerto Ricans are particularly disadvantaged.

In model 2 we consider the extent to which race and ethnic differences in high school graduation are due to differences in nativity. The effect of being foreign born on high school completion is negative and significant. The difference between Hispanics and non-Hispanic whites in high school graduation rates decreases because Hispanic groups have higher proportions of foreign

TABLE 2: Percentage Distribution for Nativity and Background Variables by Race/Ethnicity[a]

	White	Mexican	Puerto Rican	Cuban	Other Hispanic	Black	Native American	Asian
Nativity								
U.S. born	97.3	74.4	64.9	41.6	48.1	97.7	98.1	64.6
Foreign born	2.7	25.6	35.1	58.4	51.9	2.3	1.9	35.4
Foreign language								
No	91.4	8.0	4.4	3.4	17.4	98.2	95.6	33.7
Yes	8.6	92.0	95.6	96.6	82.6	1.8	4.4	66.3
Residence at age 14								
In U.S.	99.0	89.7	93.2	93.4	81.2	99.2	100.0	94.6
Not in U.S.	1.0	10.3	6.8	6.6	18.8	0.8	0.0	5.4
Sex								
Male	50.2	47.7	48.6	54.2	42.9	47.9	43.7	50.1
Female	49.8	52.3	51.4	45.8	57.1	52.1	56.3	49.9
Parental structure								
Two parents	81.0	73.1	50.0	65.1	68.0	49.8	70.2	82.2
Single parent	9.7	15.9	37.5	23.3	16.8	34.3	12.0	10.7
Stepparent	7.7	7.1	7.0	9.5	5.1	8.0	14.8	4.7
Other	1.6	3.9	5.5	2.1	10.1	7.9	3.0	2.4
Siblings								
0-1	21.5	9.0	12.5	36.2	13.3	11.7	17.0	19.6
2-3	47.5	26.6	30.7	52.0	45.5	29.8	41.4	55.1
4 or more	31.0	64.4	56.8	11.8	41.2	58.5	41.6	25.3
Parental education								
Less than high school	15.7	65.2	58.0	30.5	39.9	39.5	27.9	11.8
High school	58.4	25.9	29.5	35.6	39.3	46.3	60.4	47.4
College	24.3	3.8	7.7	29.4	17.1	8.1	9.0	35.7
Missing	1.6	5.1	4.8	4.5	3.7	6.1	2.7	5.1
N	4,288	1,047	265	96	175	2,614	343	66

[a] Percentages are calculated using sample weights.

TABLE 3: Logistic Regression of High School Graduation on Race/Ethnicity, Nativity, and Background Variables

	Model 1		Model 2		Model 3	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Race and ethnicity						
Non-Hispanic white	contrast		contrast		contrast	
Mexican	-1.139*	.075	-.981*	.079	-.408*	.130
Puerto Rican	-1.473*	.129	-1.257*	.134	-.710*	.177
Cuban	-.343	.246	.041	.255	-.189	.288
Other Hispanic	-.722*	.169	-.384*	.179	.014	.214
Black	-.593*	.059	-.598*	.059	.110	.069
Native American	-.805*	.123	-.813*	.123	-.513*	.131
Asian	-.052	.322	.210	.327	.091	.358
Nativity						
U.S. born	—		contrast		contrast	
Foreign born	—		-.637*	.093	-.280*	.115
Foreign language						
No	—		—		contrast	
Yes	—		—		.171	.120
Residence at age 14						
In U.S.	—		—		contrast	
Not in U.S.	—		—		-.992*	.173
Sex						
Male	—		—		contrast	
Female	—		—		.369*	.053
Parental structure						
Two parents	—		—		contrast	
Single parent	—		—		-.663*	.067
Stepparent	—		—		-.767*	.090
Other	—		—		-.769*	.120
Siblings						
0-1	—		—		.232*	.091
2-3	—		—		contrast	
4 or more	—		—		-.341*	.061
Parental education						
Less than high school	—		—		-.837*	.061
High school	—		—		contrast	
College	—		—		1.174*	.126
Missing	—		—		-1.242*	.124
Intercept	1.556		1.576		1.695	
-2 Log likelihood	9,632.3		9,585.8		8,688.7	
(*N* = 8,894)						

* $p < .05$ (two-tailed test)

born, and foreign birth lowers chances of graduation overall. However, Mexicans, Puerto Ricans, other Hispanics, blacks, and Native Americans are still less likely than whites to graduate from high school after controlling for nativity.

Model 3 further explains inequality in high school graduation between these groups by controlling for other relevant factors. Except for foreign language spoken in the home, all other background variables significantly affect high school graduation. Residence at the age of 14 and parental education have particularly strong effects on high school graduation. Those who resided outside the U.S. at the age of 14 had much lower chances of high school graduation, whereas those with college-educated parents had higher chances. In addition, respondents from single-parent or stepparent families and those with more siblings were less likely to complete high school. Finally, consistent with previous research, women were more likely to finish high school than men.

Controlling for these background attributes reduces the race and ethnic differentials we observed in model 2. Blacks, for example, are comparable to whites in high school completion rates once background characteristics were controlled. Furthermore, differences between whites and Mexicans, Puerto Ricans, and Native Americans diminished, but these groups still had significantly lower chances of finishing high school than whites. These findings suggest that the lower high school graduation rates of Mexicans, Puerto Ricans, and Native Americans are due to variables other than those included in model 3.

When predicting high school completion net of background characteristics, the magnitude of the foreign-born coefficient declined by more than half. Therefore, part of the negative effect of being foreign born observed in model 2 was due to the greater likelihood that foreign-born respondents had background characteristics that lowered educational attainment. These include residing outside the U.S. at the age of 14 and having parents with less than a high school education.

In analysis not shown, we examined race and ethnic differences in chances of college graduation among those who had graduated from high school. These results showed that nativity did not have a significant effect on college graduation. Thus, nativity did not help explain the lower college graduation rates of Mexicans and Puerto Ricans compared to non-Hispanic whites. Race and ethnic differences in completing college were also not due to higher proportions of respondents living in a home in which a foreign language was spoken nor to living outside the U.S. at the age of 14. What we did observe, however, was that controlling family background attributes reduced race and ethnic group inequality in college completion. Consistent with previous findings, respondents were less likely to graduate from college if they were from single-parent or stepparent families, had many siblings, or had parents with low levels of education. Because children from Puerto Rican and other Hispanic families were more likely to have these attributes, the differences between them and non-Hispanic whites disappeared when these factors were controlled. However, Mexicans, blacks, and Native Americans remained less likely than whites to graduate from college, despite the narrowing of group differences.

INTERACTION BETWEEN FOREIGN BIRTH AND RACE/ ETHNICITY

So far, our analyses have assumed that the effects of foreign birth on high school graduation are the same for each race and national origin group. To test whether the effects of foreign birth are the same for all race and ethnic groups, the analysis in Table 4 considers the interaction between foreign birth and race/ethnicity. Building on model 3 in Table 3, we added interaction variables for foreign-born Mexicans, Puerto Ricans, Cubans, other Hispanics, blacks, and Asians. We did not add an interaction variable for foreign-born Native Americans because there were too few.

The first set of coefficients in Table 4 measures the difference in high school graduation rates between the U.S. born in each race/ethnic group and U.S.-born non-Hispanic whites. Only U.S.-born Puerto Ricans and Native Americans fared worse than U.S.-born whites in chances of high school graduation, and the Puerto Rican effect was notable in that its size was twice that for Mexicans.

The coefficient for the foreign born variable in Table 4 measures the difference between U.S.-born and foreign-born non-Hispanic whites. There was no significant difference. The last set of coefficients measures the difference between the U.S. born and foreign born for each race/ethnic group. Foreign-born Mexicans were significantly less likely to graduate from high school than U.S.-born Mexicans, while there was no difference between U.S.-born and foreign-born Puerto Ricans. In a statistical test not shown, both foreign-born Mexicans and foreign-born Puerto Ricans were found to be less likely to graduate from high school than U.S.-born non-Hispanic whites.

These interactions reveal a pattern of differences in high school attainment not visible in the additive models. Net of other variables, Puerto Ricans had lower chances of completing high school than whites. Whether born on the island or mainland, Puerto Ricans fared worse than whites, suggesting a unique effect of being Puerto Rican compared to other Hispanic groups. Foreign birth clearly matters for Mexicans. Foreign-born Mexicans had lower rates of high school completion than U.S.-born whites but U.S.-born Mexicans were no different from U.S.-born whites.

There are no differences for other groups. Neither U.S.-born nor foreign-born other Hispanics had lower high school graduation rates than U.S.-born whites. The same was true for U.S.-born and foreign-born blacks, and U.S.-born and foreign-born Asians.

EFFECT OF PARENT'S PLACE OF BIRTH

In this final section, we explore the effect of parental nativity on high school graduation. Because we lack data to measure the parental nativity of Puerto Rican respondents and have small sample sizes for other race and ethnic groups, we restrict this analysis to Mexicans. Table 5 divides U.S.-born Mexicans into three groups: (1) those with two parents born in the U.S., (2) those with only one U.S.-born parent, and (3) those with two foreign-born parents. In addition, we include a category for respondents who were foreign born and had foreign-born parents and a category for respondents for whom data were missing. We contrast each Mexican group with U.S.-born non-Hispanic whites.

TABLE 4: Logistic Regression of High School Graduation on Race/Ethnicity and Nativity Interactions[a]

	Coeff.	S.E.
Non-Hispanic White	contrast	
Mexican	-.239	.142
Puerto Rican	-.645*	.206
Cuban	.254	.493
Other Hispanic	-.163	.285
Black	.094	.070
Native American	-.511*	.131
Asian	-.475	.410
Foreign born	.059	.279
Foreign born * Mexican	-.625*	.164
Foreign born * Puerto Rican	-.250	.275
Foreign born * Cuban	-.820	.575
Foreign born * Other Hispanic	.181	.367
Foreign born * Black	.644	.409
Foreign born * Native American	—	
Foreign born * Asian	1.400	.771
-2 Log likelihood = 8,668.9		
(N = 8,894)		

[a] The model controls for foreign language spoken at home, residence at age 14, sex, parental structure, number of siblings, and parental education.

* $p < .05$ (two-tailed test)

Table 5 shows that foreign-born Mexicans with foreign-born parents had significantly lower chances of high school graduation than U.S.-born whites. There is, however, diversity among U.S.-born Mexicans. Those with either one or both parents born in the U.S. have significantly lower chances of finishing high school than U.S.-born whites. Mexicans born in the U.S. to foreign-born parents, on the other hand, were more likely to graduate from high school than U.S.-born whites. We found in analysis not shown that the difference in high school graduation rates between U.S.-born Mexicans with foreign-born parents and U.S.-born Mexicans with U.S.-born parents was statistically significant. Therefore, although Table 4 documents that U.S.-born Mexicans had high school graduation rates comparable to whites, this finding masked a strong difference among U.S.-born Mexicans. Native-born Mexicans with both parents foreign born had significantly higher chances of completing high school than Mexicans with U.S.-born parents.

In analysis not shown, we examined differences between Mexican groups in their chances of graduating from college. Controlling for background factors,

TABLE 5: Logistic Regression of High School Graduation on Nativity of Respondent and Parents for Mexicans[a]

	Coeff.	S.E.
Non-Hispanic white, U.S. born	contrast	
Mexican		
Respondent: U.S. born; Parents: both U.S. born	-.344*	.153
Respondent: U.S. born; Parents: one U.S. born	-.517*	.215
Respondent: U.S. born; Parents: both foreign born	.607*	.255
Respondent: foreign born; Parents: both foreign born	-.890*	.196
Other Mexican	-.310	.306
-2 Log likelihood = 8,647.8		
(N = 8,894)		

[a] The model controls for foreign language spoken at home, residence at age 14, sex, parental structure, number of siblings, and parental education and includes variables for interactions between race/ethnicity and nativity for other race/ethnic groups.

* $p < .05$ (two-tailed test)

U.S.-born Mexicans who had two U.S.-born parents were significantly less likely to graduate from college than whites. In contrast, U.S.-born Mexicans with foreign-born parents were as likely as U.S.-born whites to finish college. The difference in chances of college graduation between U.S.-born Mexicans with two U.S.-born parents and those with two foreign-born parents was statistically significant.

Thus, parental nativity plays an important role in determining the educational attainment of U.S.-born respondents. U.S.-born Mexicans with foreign-born parents have higher high school graduation rates than either U.S.-born Mexicans with U.S.-born parents or foreign-born Mexicans. They also have higher chances of graduating from high school than non-Hispanic whites. Our analysis of college graduation showed a similar pattern. U.S.-born Mexicans with foreign-born parents were more likely to graduate from college than U.S.-born Mexicans with U.S.-born parents.

Discussion

Using data from the National Longitudinal Survey of Youth, we investigated the educational attainment of Hispanics. One characteristic which we considered may explain the lower rates of educational achievement among Mexicans and

Puerto Ricans: their greater chances of being foreign born. Because foreign birth is related to lower rates of high school graduation, we examined whether the higher proportion of foreign born among Mexicans and Puerto Ricans would explain their lower educational attainment than non-Hispanic whites. In our analysis, we found that foreign birth converted to lower rates of high school graduation. However, neither foreign birth nor other family background factors fully explained the lower educational attainment of Mexicans and Puerto Ricans.

In an analysis of interactions we discovered a more complex pattern of effects on high school attainment than those reported in the additive models. Net of other factors, foreign-born Puerto Ricans were no different from U.S.-born Puerto Ricans; both groups had lower chances of high school completion than non-Hispanic whites. Foreign birth had a strong negative effect on high school completion among Mexicans, however. Although U.S.-born Mexicans were not different from non-Hispanic whites in their chances of completing high school, foreign-born Mexicans had much lower chances.

Thus, one of our key findings is that U.S.-born Puerto Ricans have high school completion rates lower than those of U.S.-born Mexicans. Note that lower high school completion rates for U.S.-born Puerto Ricans appeared when other factors such as language, residence at the age of 14, family structure, and parental education were held constant. This is consistent with findings from a growing body of research that documents the unique disadvantage of being Puerto Rican (see Cooney 1978; Tienda 1989; Tienda, Donato & Cordero-Guzmán 1992).

Finally, by differentiating Mexicans not only by their own place of birth but also by their parents' place of birth, we found important differences among U.S.-born Mexicans. While U.S.-born Mexicans with U.S.-born parents were less likely to graduate from high school than whites, U.S.-born Mexicans with foreign-born parents were more likely to graduate from high school than whites. These results suggest that immigrant parents may pass on higher levels of motivation to their children than do native U.S. parents, and in doing so, their U.S.-born Mexican offspring have higher educational attainment than the children of U.S.-born parents.

Our analysis, therefore, has uncovered some puzzling findings related to nativity that raise questions for future research. Researchers must address (1) why place of birth does not affect the high school completion rates of Puerto Ricans; and (2) why U.S.-born Mexicans with foreign-born parents are more likely to graduate from high school than either whites or U.S.-born Mexicans with U.S.-born parents.

Answers to these questions will require examination of conditions related to the Puerto Rican experience and the educational expectations internalized by different national origin groups. For example, evidence on the residential patterns of Puerto Ricans suggests substantial segregation in the U.S. (Massey & Denton 1993). That is, similar to the experiences of African Americans in the U.S., many U.S.-born Puerto Ricans reside in places with little economic opportunity. Under such circumstances, the perceived value of a high school or college degree may be less than in other more economically successful neighborhoods, with the result that more youths drop out of school without receiving a diploma. Therefore, to develop a better understanding about the lower

achievement of U.S.-born Puerto Ricans, future models of educational inequality should consider the impact of residential segregation and neighborhood opportunities on educational attainment.

As we suggested earlier, the higher achievement of U.S.-born children of foreign-born Mexican parents may result in part from the perceptions of immigrant parents about opportunities for their children (Ogbu 1987; Rong & Grant 1992). Foreign-born parents may have higher expectations than U.S.-born parents, or they may guide their children in more ways that help promote educational attainment than U.S.-born parents. For example, school-related practices, such as parental encouragement and attention, monitoring of schoolwork, supervision of activities, and expectations, are associated with higher school achievement (Astone & McLanahan 1991). It may be that Mexican-born parents who bear children in the U.S. spend more time engaging in these practices than other parents of Mexican children.

Thus, future research of race and ethnic differences in educational achievement must emphasize both the importance of family processes and cultural perceptions of barriers to achievement. Addressing these issues requires data that measure macrolevel attributes such as neighborhood characteristics and microlevel factors such as parenting practices. We believe these may be the missing elements needed for a comprehensive explanation of the educational attainment of Mexicans and Puerto Ricans.

Notes

1. Assimilation theory predicts that the least assimilated, the foreign born, will have lower educational attainment than the partially assimilated, the U.S.-born of foreign-born parents, who in turn would have lower educational attainment than the native born of U.S.-born parents.
2. We refer to island-born Puerto Ricans as "foreign born."

References

Astone, Nan Marie, and Sara S. McLanahan. 1991. "Family Structure, Parental Practices, and High School Completion." *American Sociological Review* 56:309-20.

Bean, Frank D., and Marta Tienda. 1987. *The Hispanic Population of the United States.* Russell Sage.

Blake, Judith. 1989. *Family Size and Achievement.* University of California Press.

Center for Human Resource Research. 1991. *NLS Handbook 1991.* Columbus, Ohio: Center for Human Resource Research.

Cooney, Rosemary S. 1978. "Intercity Variations in Puerto Rican Female Labor Force Participation." *The Journal of Human Resources* 14:222-33.

Donato, Katharine M. "U.S. Policy and Mexican Migration to the United States, 1942-1992." *Social Science Quarterly* 75:705-29.

Featherman, David L., and Robert M. Hauser. 1978. *Opportunity and Change.* Academic Press.

Fernandez, Roberto M., Ronnelle Paulsen, and Marsha Hirano-Nakanishi. 1989. "Dropping Out among Hispanic Youth." *Social Science Research* 18:21-52.

Fligstein, Neil, and Roberto M. Fernandez. 1985a. "Educational Transitions of Whites and Mexican Americans." Pp. 161-92 in *Hispanics in the U.S. Economy,* edited by George J. Borjas and Marta Tienda. Academic Press.

____. 1985b. "Hispanics and Education." Pp. 113-46 in *Hispanics in the United States,* edited by Pastora San Juan Cafferty and William C. McCready. Transaction.

Haveman, Robert, Barbara L. Wolfe, and James Spaulding. 1991. "Educational Achievement and Childhood Events and Circumstances." *Demography* 28:133-58.

Kao, Grace, and Marta Tienda. 1995. "Optimism and Achievement: The Educational Performance of Immigrant Youth." *Social Science Quarterly* 76:1-19.

Krein, Sheila F., and Andrea H. Beller. 1988. "Educational Attainment of Children from Single-Parent Families: Differences by Exposure, Gender, and Race." *Demography* 25:221-34.

Lieberson, Stanley. 1980. *A Piece of the Pie: Black and White Immigrants since 1880.* University of California Press.

Mare, Robert D. 1979. "Social Background Composition and Educational Growth." *Demography* 16:55-71.

____. 1981. "Change and Stability in Educational Stratification." *American Sociological Review* 46:72-87.

Massey, Douglas S., and Nancy Denton. 1993. *American Apartheid: Segregation and the Making of the Underclass.* Harvard University Press.

Massey, Douglas S., Rafael Alarcón, Jorge Durand, and Humberto Conzález. 1987. *Return to Aztlan: The Social Process of International Migration from Western Mexico.* University of California Press.

McLanahan, Sara. 1985. "Family Structure and the Reproduction of Poverty." *American Journal of Sociology* 90:873-901.

Ogbu, Robert U. 1987. "Variability in Minority School Performance: A Problem in Search of an Explanation." *Anthropology and Education* 18:313-34.

Portes, Alejandro, and Ruben G. Rumbaut. 1990. *Immigrant America: A Portrait.* University of California Press.

Powell, Brian, and Lala Carr Steelman. 1990. "Beyond Sibship Size: Sibling Density, Sex Composition, and Educational Outcomes." *Social Forces* 69:181-206.

Rong, Xue Lan, and Linda Grant. 1992. "Ethnicity, Generation, and School Attainment of Asians, Hispanics, and Non-Hispanic Whites." *Sociological Quarterly* 33:625-36.

Rumberger, Russell W. 1983. "Dropping Out of High School: The Influence of Race, Sex, and Family Background." *American Educational Research Journal* 20:199-220.

Sandefur, Gary D., Sara S. McLanahan, and Roger A. Wojtkiewicz. 1992. "The Effects of Parental Marital Status during Adolescence on High School Graduation." *Social Forces* 71:999-1017.

Tienda, Marta. 1989. "Puerto Ricans and the Underclass Debate." *Annals of the American Academy of Political and Social Science* 501:105-19.

Tienda, Marta, Katharine M. Donato, and Hector Cordero-Guzmán. 1992. "Schooling, Color, and the Labor Force Activity of Women." *Social Forces* 71:365-98.

U.S. Department of Education. 1992. "Are Hispanic Dropout Rates Related to Migration?" National Center for Education Statistics, Government Printing Office.

Velez, William. 1989. "High School Attrition among Hispanic and Non-Hispanic White Youths." *Sociology of Education* 62:119-33.

Wojtkiewicz, Roger A. 1993. "Simplicity and Complexity in the Effects of Parental Structure on High School Graduation." *Demography* 30:701-18.

Economic Decline, Gender, and Labor Flexibility in Family-Based Enterprises: Midwestern Farming in the 1980s

LINDA LOBAO, *Ohio State University*
KATHERINE MEYER, *Ohio State University*

Abstract

During economic downturns, traditional gender allocations of labor have been considered to vary more than in prosperous times. While most studies have examined the division of labor in the household or in paid employment, we examined it where both intersect, in family-owned and family-operated enterprises in the farm sector of the 1980s. This context, combining crisis conditions and the agency of economic actors, should be related to greater flexibility in labor allocations, leading to the feminization of farming. However, a contrasting perspective argues for rigidity of gender roles in farming. We use data from a twelve-state midwestern sample and a more detailed Ohio study. The results failed to support the flexibility thesis. The rigidity of production roles was further translated into different factors related to women's and men's stress.

Gender divisions of labor usually are considered most malleable during periods of economic decline (Hareven 1990; Lipman-Blumen 1984). In such times, survival necessitates rationalistic allocation of labor resources, presenting openings for women to broaden participation in work and occupations formerly reserved for men. Most research on the division of labor in the contemporary U.S. focuses on separate spheres of paid employment in the formal sector and household reproductive activities. This study examines work roles where household and enterprise intersect, in family-owned and family-operated businesses. We focus on an economic sector that experienced a dramatic downturn in the 1980s — midwestern farming — and examine gendered

* *This research reports on two projects supported by grants from the following institutions: the National Science Foundation under grant #SES-8809467; the North Central Regional Center for Rural Development; the U.S. Department of Agriculture, Economic Research Service; the Ohio State Seed Grant Program; and the Ohio Agricultural Research and Development Center. We thank William Form, Toby Parcel, Barbara Reskin, and anonymous reviewers for their comments; and Maurice Klein for providing economic trend data. Direct correspondence to Linda Lobao, Rural Sociology Program, or to Katherine Meyer, Department of Sociology, The Ohio State University, Columbus, OH 43210.*

patterns of work in the farm enterprise, household, and paid employment during the farm crisis. In particular, we question whether a traditionally male-dominated industry became more feminized and document the implications of gendered work patterns for emotional well-being.

Several literatures address the gender allocation of labor in times of hardship and economic decline. Historical work examines women's employment during war and depression (Elder & Liker 1982; Hareven 1990). Studies of Third World women document their use of work-related survival strategies, including informal sector employment, to cope with economic hardship from national structural adjustment (Daines & Seddon 1993; Moser 1992). Research on contemporary Western societies describes how gendered patterns of paid employment vary by industrial sector during recessionary periods (Bagguley et al. 1990; Cummings 1987). These literatures view macrolevel changes as the structural context of a historical period. Change washes over economic actors differentially, depending upon prior social statuses. Women's domestic responsibilities are argued to remain constant despite added productive work outside the household. As a consequence, some analysts contend that women bear the costs of economic crises, in terms of multiple work burdens and higher stress relative to men. The farm sector during the 1980s provides a useful context for elaborating these employment and personal outcomes.

Most knowledge about contemporary economic structure is from research on paid employment, large firms, and urban settings. Less is known about family enterprises that have (1) no clear separation between labor and capital, (2) depend mainly on unwaged labor, (3) integrate spheres of production and consumption, and (4) share other informal sector attributes (Aronson 1991). The survival of such enterprises in advanced economies are of increased sociological concern (Castells & Portes 1989; Collins 1990; Mingione 1991). The ability to mobilize unpaid household labor, particularly that of women, results in a more flexible work force, enhances competitiveness, and enables small-business owners to resist taking salaried employment. The advantages of household labor have been well-documented in the case of farming and explain why family businesses dominate numerically in that sector (Friedmann 1978). Highly fluctuating market conditions have characterized farming historically. During downturns, households are able to sacrifice economically by increasing farm and off-farm work and by reducing consumption levels. While such strategies involve all family members, women have been seen as particularly stressed and triply burdened with household work, farm work, and nonfarm income-generating activities.

Farming is also an industry where the division of labor historically has revolved around gender (Rosenfeld 1985). It provides an opportunity to examine whether traditional divisions of labor relaxed during crisis. More broadly, as family-based production, farming enables the exploration of how producers themselves, in settings less constrained by capitalist managerial imperatives, act to preserve or eliminate gender-segregated activities. Farming during the 1980s thus offers a context in which crisis conditions and the flexibility of economic actors should alter traditional work roles. Our focus is to examine employment and personal outcomes in this context of general economic decline and through the farm household variations within.

Context of Farming in the 1980s: The Historical Period and Farm-household Differentiation

The 1980s represents the worst financial stress in the farm industry since the Depression (Lasley et al. 1995; Murdock & Leistritz 1988; Stam et al. 1991). Some analysts view the crisis as a fundamental breakdown in the regulatory structure of farming under Fordism, particularly due to changes in globalized agricultural commodity markets (Kenney et al. 1989). While other interpretations have been posited, most scholarly work recognizes that the crisis was the result of both long-term structural trends and specific economic events. Historically, U.S. agriculture suffered from problems of overproduction, unstable and low farm incomes, and high government support costs. Exportation of agricultural commodities was encouraged in the late 1960s as a strategy for managing these problems. In the process, farming became more vulnerable to new sources of price and income instability (de Janvry & LeVeen 1986). Over the 1970s, farm exports expanded almost sixfold. Farm prices rose, and farmers expanded operations by investing in land, machinery, and equipment. Often they borrowed heavily to do so (Leistritz & Murdock 1988).

The conditions fostering the agricultural boom of the 1970s reversed in the early 1980s. Farmers were faced with a national recession and a decrease in world demand for U.S. products due to the rising value of the dollar, the Soviet grain embargo, and increased market competition from other producer nations. Interest rates and the costs of producing farm commodities continued to rise and commodity prices fell. As returns to farmland diminished, land values declined: they fell 27% nationally from their 1981 peak to almost 60% in 1986 in export-sensitive midwestern states (Leistritz & Murdock 1988). Real median real estate value plunged from $351,157 per farm in 1982 to $270,471 in 1987, its lowest rate of the crisis (Jones & Canning 1993). Declining land and other asset values eroded farmers' equity and, coupled with lower farm income, made servicing debt difficult. By 1986, about one-fifth of all farmers were considered to be in financial stress. Forced sales of farmland jumped from 19% of all farmland transfers in 1980 to 46% in 1986. The percentage of farms going out of business tripled and the percentage going out through bankruptcy more than quadrupled from 1982 to 1986. By 1990, from 8% to 12% (200,000-300,000) of those farming in 1980 failed financially, becoming bankrupt, foreclosed, or financially restructured (Stam et al. 1991:2).

We should note that analysts refer to the farm crisis as beginning in the early 1980s rather than linking it to any one year or specific event. Signs of crisis are documented retrospectively, with the years 1981-83 representing clear shifts toward rising real interest rates, declines in land values, and plummeting farm exports. However, the severity of these changes for the sector's financial well-being was not recognized by the federal government or academia until about 1984, mainly because of the lack of current information about farm debt distribution, bankruptcy, and other financial conditions (Harl 1990; Stam et al. 1991). The media introduced the farm crisis to the public mid-decade (Harl 1990).

While the farm crisis represented an industry-wide decline, specific enterprises were affected differentially, as in any sectoral downturn. The crisis

varied across regions, enterprise size, and commodity specialization. Midwestern farms were hard hit due to their cash grain specialization. Commercial farms defaulting on loans were concentrated in this region (Hanson 1990). Financial stress affected particularly middle- to upper-middle-size commercial units (about $100,000-$250,000 in gross annual sales) that are less able to manage high debt loads. Younger farmers were affected because they were newer, more highly leveraged entrants into farming. In the case of family enterprises, declines tend to be directly reflected in household well-being. Economists have long recognized the fungibility of financial assets in family enterprises (Aronson 1991), and sociologists have stressed parallel linkages of production and consumption (Mingione 1991). During the crisis, farm households were observed to cutback on consumption expenditures in accordance with enterprise conditions (Barlett 1993; Belyea & Lobao 1990; Johnson, Lasley & Kettner 1991). The household conditions of the farm population thus varied during the crisis period.

A characterization of the 1980s farm economy must consider not only the financial crisis but also previous rounds of restructuring that differentiated enterprises and households. Paralleling the development of other industries in postwar Fordism, farming grew more capital-intensive and concentrated, an increasingly specialized mass producer of food. This system, which emerged in part from New Deal policies to alleviate the Depression in farming, performed quite well until the crisis period, albeit with enormous consequences for farmers and rural people. From 1940 to 1990, the number of farms declined by two-thirds and the farm population shrank from nearly a quarter of all Americans to about 2%. Most of this massive restructuring had occurred by the 1970s. A dualistic farm system now exists marked by a relatively large number of small, part-time farms and a few large farms with expanding market-shares. Moderate-size family farms, those closest to the petty bourgeois ideal in which household members own the major production factors, use minimal hired labor and depend mainly upon farming for livelihood are being edged out of this system (Krause 1987; Lobao 1990). Concomitantly, the class structure of the farm population has become increasingly internally stratified (Mooney 1988). Producers occupy a range of class positions on both sides of the petty bourgeoisie, from semi-proletarianized to capitalist or hired-labor-dependent farmers. In summary, the 1980's farm crisis was an industry-wide, cyclical downturn that occurred in a system in which enterprises varied in size, in asset holdings, and in general production-related characteristics. Households varied by consumption conditions and class location as well.

Gender Allocations of Labor in a Context of Decline: Flexibility vs. Rigidity

Two contrasting perspectives on the broadening of work in direct production can be delineated. Both assume that in family enterprises, production and consumption and farm and household well-being are linked. Both see individuals' labor allocations as a function of production and household characteristics, but they differ in including gender as a guide in this process. Both literatures treat relationships as period and position effects. That is, they view women's

and men's labor allocations as adaptations to a particular economic environment. Labor allocations vary among family enterprises to the degree that they are marginalized in this environment.[1]

FLEXIBLE LABOR ALLOCATIONS: OFF-FARM WORK AND THE FEMINIZATION OF FARMING

The first perspective suggests that both women's and men's participation in production should be more flexible in periods of downturn. Enterprise stress confers shared household conditions and experiences. Farm men and women should adapt by intensifying or expanding production and/or increasing off-farm employment. Relative to the past, farm women are likely to have greater involvement in farming and off-farm employment in the 1980s. Several literatures support this expectation of work role broadening.

Neoclassical economic theory, as articulated in the new home economics, assumes that households are decision-making units. Household welfare is dependent upon rational allocation of labor and consumption resources and on adaptation to economic environment (Becker 1981; Pollak & Wachter 1975). In order to maximize welfare, households should allocate members' home and market work time by criteria such as human capital skills, external labor market characteristics, and substitutability of purchased home inputs for member-provided ones. While the new home economics has been criticized by sociologists on a variety of fronts (England & Kilbourne 1990; Spitze 1986), it has served to bring household dynamics into the analysis of the farm enterprise (Huffman 1991). Nearly all research on farm household labor by economists or home economists is framed in terms of neoclassical assumptions. Most studies focus on the operator's and, less frequently, on spouse's off-farm employment and were conducted prior to the farm crisis. Virtually all research is cross-sectional. From this perspective, households allocate men and women to farming or nonfarm employment based upon household, farm enterprise, and local labor market exigencies (Findeis, Lass & Hallberg 1991). Gender segregation in farming activities due to ideology, tastes for tradition, or patriarchal control is little addressed but would be seen as nonrational and diminishing in times of crisis.[2]

Political economy perspectives also have emphasized the flexibility of family labor in periods of hardship and change (Walby & Bagguley 1989). While conflicting with neoclassical economic perspectives about why family enterprises survive under late capitalism, this literature often makes similar assumptions about household consensus and the rationality of survival strategies. Maintaining the enterprises' economic position, particularly so that wealth can be passed down to subsequent generations, tends to be viewed as equally important for all household members. In effect, class-based group interests are seen to override potential gender interests in allocating household labor. As noted above, historical evidence shows that women's labor in farming and in nonfarm employment has been mobilized during economic downturns, enabling family farmers to survive and compete with larger-scale units (Friedmann 1978; Stock 1992). However, the extent to which this has occurred in recent years is unclear.

While there is little quantitative information on shifts in women's farm work during the 1980's farm crisis, researchers have extrapolated from case

studies and earlier work to contend that farming became more feminized. Gladwin (1991) argues that role flexibility likely increased during the crisis. Her north Florida study found that women reported spending more time at farming and were a labor substitute for their spouses who were working off-farm to maintain family income. Such substitution occurred in part because rural men have higher wages than women. Friedland (1991) notes that general restructuring of farming in the 1980s, particularly the relative increase of part-time farming on small farms, did feminize farm work. Other work that does not specifically address the U.S. crisis supports these contentions. Some studies that focus on the structural position of the enterprise have found that women participate more in production on smaller, less-profitable units (Buttel & Gillespie 1984; Coughenour & Swanson 1983); and conversely, that gender segregation in farm work and marginalization of women from decision making is greater on higher-volume commercialized units (Sachs 1983). Studies of other first world countries also suggest that production is feminized when units are a less viable means of household support (Pfeffer 1989).

Finally, literature on occupational segregation, although centered on paid, formal sector employment, may offer insights about labor allocation in family enterprises. Arguments that women constitute a "reserve army of labor" when the supply of male labor is low (Jones & Rosenfeld 1989) support previous contentions that women perform farm labor for spouses who work off-farm or to reduce the costs of hired help. Furthermore, employment demand for men may be higher because rural labor markets are often composed of extractive industries and because the midwestern markets are composed of durable manufacturing. Research indicating women's greater representation in sectors that have been deskilled and poorly remunerated (Tickamyer & Bokemeier 1988; Walby & Bagguley 1989) also suggests that women's presence in farming should be greater during downturns.

RIGIDITY IN FARM LABOR ALLOCATIONS

A contrasting perspective is from feminist and other literature that focus on statuses within the family. Although this literature suggests that both men and women allocate labor flexibly outside the family enterprise, the reasoning for this is grounded in gender-based rather than collective household interests and experiences. Moreover, it indicates that the division of labor on the enterprise itself will be little altered in contexts of decline. Edwards (1979) characterizes small enterprises by their use of coordination or control systems. In the former, work is organized by consensus, through workers' recognition of mutual production goals, shared control or ownership over factors of production, and equitable remuneration of work. In small enterprises using simple control, positions in production take a more hierarchical form, with a designated person or persons directing, evaluating, rewarding, and disciplining the work of relatives and hired labor. Literature that centers on collective household interests in the allocation of labor by family businesses implicitly assumes the use of coordination over control. In farming, as in other family businesses, both processes operate to varying degrees. Methods of controlling and structuring the

labor force penetrate family farming in the guise of kinship relations of gender and generation (Wilson, Simpson & Landerman 1994).

Recent feminist work has documented the intransigence of asymmetrical family relationships in farming. Most women confront farming through their positions as wives, in structurally subordinate positions in the family enterprise (Whatmore 1991). This position confers specific roles and responsibilities, subject to human agency, in farm production and household reproduction. Women are situated in a "patriarchal labor process" in which the timing, nature, and outcomes of their farm work tend to be directed by men and coordinated around household exigencies. While the gender division of farm labor varies by enterprise characteristics, commodity, family norms, ethnicity, and other factors, the general trend is that men retain greater control over operating decisions such as marketing, input purchases, and major production tasks such as field work (Barlett 1993; Rosenfeld 1985; Salamon 1992). Men's control is reinforced by their greater property ownership and by institutions such as the Agricultural Extension Service, agricultural cooperatives, and farm organizations with a legacy of serving men's production needs foremost (Sachs 1983). As a consequence, men stand at the apex of the operation and women in more distant roles (Simpson, Wilson & Young 1988). Thus, the distinct statuses of men and women in the family enterprise suggest that women should experience and respond to downturns in different ways from men. From this perspective, traditional divisions of labor are likely to remain fixed in crisis.

Evidence supporting rigidity in farm work is from studies finding that men's and women's labor allocations do not revolve similarly around enterprise exigencies, which distinctly influence men's farm and off-farm labor but tend to have little effect on women's (Simpson, Wilson & Young 1988). The greater the commercialization or scale of the farm and the more hired labor use, the more time men allocate to farm work and the more they reduce off-farm work, relationships that do not hold for women. In contrast to studies above that find farm work more feminized on smaller, less-profitable units, other work points to a continued lack of women's substitution for men and hired labor (Whatmore 1991; Wilson, Simpson & Landerman 1994). Enterprise commercialization is shown to be little related to women's work in both farming and nonfarm employment (Simpson, Wilson & Young 1988) or to affect women's farm work only in certain ethnic and community contexts (Salamon 1992; Shaver & Reimer 1991). These gender differences are reflected in the determinants of farm and off-farm work, age, education, and other demographic characteristics for both women and men and enterprise characteristics for men (Lyson 1985).

In sum, this perspective emphasizes that women are situated in the farm-household production nexus in different ways from men. The experiences and interests associated with this position should structure their response to crisis differentially as well. Because women historically have had blocked opportunity in farm management, their likely response to perceived hardship and change is to look first to the nonfarm economy. Hence, women's labor allocations are thus likely to remain rigid in the farm enterprise but elastic with regard to off-farm work.

Production, Household Reproduction, and Stress: The Gendered Costs of Labor Allocations

Focusing on labor allocations in the crisis period allows us to elaborate upon a frequently echoed theme, that the costs of economic crises are gendered, with women often experiencing greater hardship than men in terms of increased work burdens and personal stress (Daines & Seddon 1993; Elder & Liker 1982). Women's stress in hardship contexts has been attributed to their continued responsibility for household reproduction despite increased work outside the household and to women's greater sensitivity to declines in consumption levels, resulting from their roles as household and community managers (Coverman 1989; Hochschild 1989; Tilly 1981). Farm women's stress was recognized as a national social problem at the turn of the century, as the Roosevelt administration feared that women's extensive work burdens would spur farm families toward rural out-migration and, in turn, escalate food costs (Sachs 1983). Some studies report that contemporary farm women have significantly higher levels of social-psychological stress than men (Meyer & Lobao 1994; Walker & Walker 1988; Duncan, Volk & Lewis 1988), attributing this to the multiple roles these women play and to their day-to-day management of household consumption. More detailed case studies of farmers during the crisis link social-psychological stress to household economic position and labor allocations in different ways for men and women (Barlett 1993; Conger & Elder 1994; Rosenblatt 1990). For men, stress is more connected to their occupational status as farmers while for women it is more connected to family lifestyle. We extend our analysis to women's and men's stress to assess the extra-economic costs of pursuing gendered strategies of labor allocations and to examine the implications for farm peoples's ability to underwrite the present agricultural system.[3]

Research Questions

Although a number of studies examine farm women's and men's farm- and off-employment, few consider the 1980's crisis. Of those existing, nearly all focus on off-farm employment. Farm women, relative to farm men, to other rural women, and to the past, appear to have increased their participation in the labor force in the 1980s (Ollenburger, Grana & Moore 1989). McDonald and Peters (1991) examined the labor force participation of farm women and men in seven West North Central states using data from the *Current Population Survey*. In 1989, 58% of farm women 16 and older were in the labor force as compared to 45% in 1979. A number of questions remain about farm household labor allocations during the crisis period. First, the extent to which farm enterprise and household conditions influenced off-farm labor allocations is unclear. Studies using secondary data, such as those above in which farmers are a subset of the general population, are limited because they lack detailed enterprise data such as debt levels, enterprise capitalization, commodity specialization, and household indicators of financial stress. As a consequence, researchers have inferred rather than directly demonstrated that women's off-farm participation increased due to the hardships of the crisis. Second, as noted above, there is

little information about whether the traditional division of labor in the enterprise itself was altered. Third, the household division of labor is generally unexplored. Although studies of other populations indicate that women's domestic responsibilities remain unchanged, some case-study accounts suggest that men increased domestic tasks as wives took on greater roles in farm or off-farm production (Davidson 1990).

We analyze these questions by considering the farm crisis in two ways, as a structural or macrolevel context and through the gradations of household-enterprise economic positions within this context. As noted earlier, analysts have typically analyzed decline in terms of temporal setting and/or through differential positions that manifest decline. Based upon the previous literature, we expect that in the 1980s, expansion of work domains to the nonfarm sector is likely for both men and women. Total time spent in productive work may also increase. Men may intensify their on-farm work, particularly those with larger operations. Perspectives about women's farm work lead to contrasting hypotheses: one indicates a greater feminization of production, the other indicates little change in women's farm activities and even possible decrease as women exit farming for off-farm employment. Within this context, differential household enterprise characteristics should shape labor allocations. In family businesses, production and consumption are intermixed so that labor allocations of both men and women should be related to enterprise and household financial stress. However, the two perspectives differ as to how gender guides labor allocations and its outcomes. One implies increasing similarity between men and women in interests and experiences of crisis and substitutability of women's for men's labor. The other argues that because women occupy differential structural positions in the family enterprise, their interests and experiences likely vary from men's, their labor allocations stand to be affected by different factors, and they likely will not substitute for men working off-farm or on marginal units. Finally, we pose questions about outcomes beyond the allocation of work in production. Did men increase their share of domestic duties? To what extent did changes in work patterns and related crisis experiences result in differential patterns of stress for women and men?

Data and Methods

This study is based on data from two sources. Considering the crisis period as a structural context within which farm people acted, we present a broad description of the patterns and reported changes in women's and men's work activities for twelve midwestern states. We used a more detailed Ohio sample to examine the gradations of household-enterprise economic conditions within this context and the extent to which these were related to women's and men's labor allocations.

THE MIDWEST STUDY

The midwest is composed of the Lake states (Minnesota, Michigan, and Wisconsin), the Corn Belt (Ohio, Illinois, Indiana, Iowa, and Missouri), and the Plains states (North Dakota, South Dakota, Kansas, and Nebraska). In 1988, about one-fifth of Midwest farmers had debt-to-asset ratios (a commonly used measure of financial stress) over the standard 40%, indicating serious-to-extreme financial difficulty. This proportion of operators was nearly double that of the south and southeast and higher than all other regions (Morehart, Johnson & Banker 1988). Between 1980 and 1990, 64% of all counties in the region lost population, a reflection of the farm crisis and downturns in other rural industries.

In 1989, data were collected from a random sample of farm households within each of the twelve states. Two questionnaires were sent — one for the "person responsible for the majority of farm operation decisions" and one for his/her spouse or household partner. Single farmers were asked to complete only one survey, which focused mainly on production needs and techniques and farm-household finances. The partner survey focused on changes in the division of labor and personal involvement in farming as a result of the crisis, perceptions about adjustments, and life stressors. A total of 4,087 individuals completed the production survey, and their 3,630 respective spouses or household partners completed the partner survey.[4] About 98% of those completing the partner questionnaire were women and 97.3% of those completing the production survey were men. About one-quarter of the farms had debt-to-asset ratios exceeding 40%. About 42% of the sample had net family incomes of $20,000 and below.

THE OHIO STUDY

The Ohio study provides more detailed comparisons of men and women. The Ohio study is part of a longitudinal project monitoring approximately 900 farms selected randomly from a list of all Ohio farms (approximately 70,000) that were provided by the Ohio Agricultural Statistical Service. Telephone interviews collecting basic farm production and household demographic information were conducted in 1988. A detailed financial balance sheet was created for income and expenses for the previous year, 1987. Separate mail-out questionnaires focusing on the division of farm and household labor (including the major variables used in this study) were then sent to each operator and spouse/household partner. This analysis uses data from the 531 men and 497 women who completed the mail-out survey and whose farms were included in the telephone interview. About 98% of the women and 90% of the men are married and the sample was composed of 86% matched couples. Farm characteristics typify the Corn Belt pattern of smaller commercial family farms. Most farmland is owned by the household, and enterprises use little hired labor. The pattern of smaller farms results in a less highly leveraged population than in the Plains or Lake states of the midwest. The Ohio sample was similar to the rest of the Cornbelt in that about 17% of farmers had debt-to-asset ratios exceeding 40%, above the national average. Median family income in 1987 was $26,331 for the sample, about $4,500 below that for the nation for the same year

(U.S. Bureau of the Census 1991). About the same proportion (43%) as in the midwest farm survey had family incomes lower than $20,000.

The Gender Division of Labor and Work Patterns during the Crisis Period

WOMEN'S AND MEN'S INVOLVEMENT IN ENTERPRISE, HOUSEHOLD, AND OFF-FARM WORK

Women were asked to report how often they perform a variety of tasks that take place on their farm following Rosenfeld's (1985) delineation, and whether they do these regularly, occasionally, or never. These data are shown for women completing the partner survey for the midwest ($N = 3{,}391$) and for both men and women in Ohio (Table 1). Although there are three farm production subregions within the midwest, there is little regional variation among them in any of the following analysis, making it more parsimonious to report regional totals. Women's involvement in direct farm production is similar to that reported in earlier studies. Most perform farm tasks at least occasionally but are at a greater distance from management responsibilities (Rosenfeld 1985; Simpson et al. 1988). Although women did bookkeeping on a regular basis, they were far less likely to do field work and to participate in management tasks such as marketing, purchasing inputs, and supervising labor. The Ohio data revealed gender differences more distinctly, as data on division of labor were collected for both men and women. More than half the women contributed at least occasionally to harvesting crops, raising farm animals, running errands, and keeping a home garden. At the same time, they were nearly wholly responsible for household tasks.

The frequency of off-farm employment across the midwest and in Ohio is reported in Table 2. Off-farm employment refers to whether or not the respondent worked at either a paid job outside the home or farm and/or was self-employed (in addition to farming).[5] Women were more likely to be employed off-farm than men. Respondents were also asked whether or not they took off-farm employment "because of financial need in the past five years." Since the widespread realization of the crisis circa 1984, 38% of women took off-farm jobs compared to 23% of men. Thus, farm people reported further semi-proletarianization and at rates that varied by gender. As would be expected from the postwar differentiation of the farm sector, however, most people working off-farm did so prior to the crisis with men having worked a median of 16 years and women, 9 years. About two-thirds of midwestern farm women and three-quarters of men worked off-farm 30 hours or more per week during the crisis (Table 2). Women seemed to occupy jobs created during different epochs of rural economic restructuring. About one-third were in traditional women's rural professional jobs, such as teaching and nursing. Historically, education and hospitals and health services have been large employers in rural areas. Nearly half of the women occupied lower-wage jobs in the more recently expanding service sector. The more detailed Ohio sample shows that about 20% of employed women were self-employed and engaged in informal sector activities such as baby-sitting, house cleaning, and providing business services. About one-third of midwestern net family income came from women's

TABLE 1: Labor Allocations to Farm Household Tasks by Gender[a]

	Midwest		Ohio			
	Women		Women		Men	
Percent reporting task[b]	Regular Duty	Occas. Duty	Regular Duty	Occas. Duty	Regular Duty	Occas. Duty
Plowing, disking, cultivating, or planting			6	37	82	15
Applying fertilizers, herbicides, or insecticides			1	6	69	26
Doing other field work without machinery			3	40	51	43
Harvesting crops or other products including running machinery or trucks[c]	11	57	10	41	83	14
Taking care of farm animals, including herding or milking dairy cattle	25	50	23	38	79	15
Running farm errands, such as picking up repair parts or supplies	29	66	28	62	69	30
Making major purchases of farm supplies and equipment	5	28	6	28	88	11
Marketing your products, dealing with wholesale buyers or selling directly to consumers	6	23	6	21	79	15
Bookkeeping, maintaining records, or preparing tax forms for the operation	53	29	51	26	58	32
Doing household tasks like preparing meals, housecleaning[d]	94	5	98	2	7	57
Supervising the farm work of other family members[e]	5	41	10	39	48	41
Supervising the farm work of hired farm labor			4	30	58	28
Taking care of a vegetable garden or animals for family consumption	60	32	58	29	37	50
Looking after children			70	19	5	71
N	3,391		497		531	

[a] Follows Rosenfeld's (1985) classification of farm tasks.
[b] Total excludes those who report that the task was "not done" on their operation.
[c] Summary question of first four items reflecting general field work for midwest only.
[d] Combines household tasks and child care for midwest only.
[e] Combines family and hired workers for midwest only.

TABLE 2: Participation in Off-Farm Employment by Gender

	Midwest Region		Ohio	
	Women	Men	Women	Men
	(Percent)		(Percent)	
Any off-farm employment	50	40	49	42
Percent of those employed who took employment during crisis period	38	23	—[a]	
Hours per week				
1-9	7	6	9	10
10-19	10	9	12	6
20-29	16	8	17	8
30-39	16	8	24	8
40+	50	68	38	68

[a] Not asked for Ohio

employment, slightly higher than the proportion contributed by working wives nationally (Ritchie & Pollard 1994).

CHANGES IN LABOR ALLOCATIONS DURING THE CRISIS PERIOD

Midwest women were asked to report changes in work roles in the past five years or since about 1984 (Table 3). Most reported that the time they spent at various farm and household tasks stayed the same. For those reporting changes, declines exceeded increases for all categories of farm work except bookkeeping. The Ohio sample showed specific gender differences. A greater proportion of men increased rather than decreased time spent in caring for farm animals, marketing, and bookkeeping. Men's reported decreases were in supervising labor and in field work but those changes were lower than those of women. These differences are reflected in changes in general farm involvement, with a higher proportion of men reporting increased relative to decreased involvement in farming and women reporting the opposite.

The amount of time spent at an off-farm job increased for 20% and decreased for 16% of the women. In Ohio, 28% of women reported increases as compared to 17% of men and only 8% of both men and women reported decreased off-farm work. In Ohio, off-farm employment appears to have accelerated during the crisis period, possibly due to the greater availability of nonfarm jobs and to the smaller size of farm operations. In contrast to case-study observations, men did not increase their share of domestic work relative to the past or to women. While changes in work involvement in farming and nonfarm employment may have reflected the effects of aging, eliminating farmers older than 70 showed the same gender gap and relatively little change in the percentages reported in

TABLE 3: Changes in the Amount of Time Spent on Farm, Household, and Off-Farm Work since 1984

	Midwest		Ohio			
	Women		Women		Men	
	Increase (Percent)	Decrease	Increase (Percent)	Decrease	Increase (Percent)	Decrease
Field work	14	31	7	33	18	23
Care of farm animals	15	31	11	22	20	18
Marketing products	6	13	4	13	17	13
Bookkeeping	25	10	16	8	23	8
Household tasks	20	14	15	12	4	16
Supervising farm work	6	16	2	19	9	21
General farm involvement[a]			14	28	19	13
Time spent in off-farm job	20	16	28	8	17	8

[a] Not asked for midwest sample.

Table 3. In sum, there was little evidence that farming became feminized during the crisis period. Rather, for women who changed work roles, the most likely path was to become more distant from farming, relative to men, by expanding off-farm employment.

Household-enterprise Position during the Crisis: Gendered Labor Allocations and Stress

The Ohio study provided a more detailed, empirical test of the factors affecting the gender allocation of labor in paid work and in farming. We hypothesized that the financial position of the household-enterprise during the crisis is related to the labor allocations of farm people.[6] Further, due to men's and women's respective structural positions in the farm household, we expected that different factors influenced their labor allocations. Women also should be more sensitive to economic restructuring, as it is manifested through household financial hardships, and possibly experience greater social-psychological stress.

MEASURES

All variables are described below. Means and significant differences between men and women based on *t* tests are shown in Table 4. We employed five measures of work involvement. Whether or not the respondent participated in *off-farm work* and *off-farm work hours,* were described previously. For the Ohio sample, off-farm work and off-farm hours specifically included time spent in nonfarm informal sector activities, such as home-based production and personal services. We reported off-farm work hours only for those who had off-farm work. Time spent at *farm work* was calculated from the average number of hours spent weekly for each season on farm work (however the respondent defined

TABLE 4: Means for Variables by Gender for Ohio

	Mean		Results of Differences of Means Tests
	Women	Men	
Labor allocations			
Off-farm work	0.49	0.39	**
Off-farm work hours	31.36	35.86	**
Changes in off-farm work time	2.21	2.08	***
Farm work (hours)	12.30	36.13	***
Changes in farm-work time	1.86	2.09	***
Farm structure			
Gross farm sales	3.65	3.57	n.s.
Hired labor use	0.60	0.59	n.s.
Cash grains	0.40	0.40	n.s.
Dairying	0.22	0.20	n.s.
Farm-household financial conditions			
Debts-to-assets	0.20	0.18	n.s.
Financial adjustments	0.34	0.30	*
Perceived economic hardship	4.39	4.16	**
Dissatisfaction with finances	8.50	7.93	**
Family income	$33,017	$31,775	n.s.
Sociodemographic characteristics			
Ownership of capital assets	2.42	3.43	***
Young children	0.31	0.27	n.s.
Farm background	1.55	1.89	***
Age	49.51	52.66	***
Education	4.67	4.41	***
Spouse works off-farm	0.39	0.35	n.s.
Labor market context			
Metropolitan proximity	2.32	2.33	n.s.
Percent of labor force that is female	0.38	0.39	n.s.
Role stress	12.70	12.22	*

* $p < .05$ ** $p < .01$ *** $p < .001$ n.s. = Not significant

this). Respondents indicated their work time on a five point scale of hourly intervals for all four seasons. The midpoint of each hourly interval was taken and then divided by 4. Men reported spending about 36 hours per week annually in farming and women, about 12. *Changes in off-farm and farm work time* referred to the last two items in Table 3, whether this work increased (3), stayed the same (2), or decreased (1) since the crisis.

We delineated independent variables that reflect structural attributes of the enterprise, enterprise-household conditions, and control variables for personal/sociodemographic characteristics and labor market context that are commonly used and central to the two perspectives above on farmers' labor allocations.

Farm Enterprise Structure

Enterprises varied as to their structural position in the farm industry at the time of the crisis. We used several measures of farm structure that have been commonly associated with labor allocations. Extent of farm commercialization is measured in terms of size and hired labor use. Farm size is measured through a commonly used measure, *gross sales* from farming. Farms were placed into one of seven standard size classes ranging from (1) under $10,000 to (7) $500,000 and over, with median sales about $40,000. Because most farms used little hired labor (with total median annual usage about 80 hours), we created a dichotomous variable, coded 0 for no hired labor and 1 for use of hired labor. The commodity produced affects the amount of time that must be spent on farm work and is linked to prices received during the crisis. Farms were classified by two types of commodities, cash grains and dairying. Cash grain operations require less continuous labor. Cash grains also experienced their lowest prices of the 1980s in 1986-87. Dairy operations are controlled because they require continuous labor, constraining off-farm employment. *Cash grain dependency* is measured by the proportion of gross farm income from sales of cash grain and inventory value. Farm operations are coded 1 if mainly *dairying,* 0 if not. Table 4 shows no significant differences between men and women on the farm enterprise variables as would be expected since most of the sample are matched couples.

Farm-Household Financial Conditions

The financial position of household and enterprise are intertwined in family enterprises. We used both objective and subjective or perceptual indicators of financial conditions to indicate varying positions among farm households during the crisis. The *debt-to-asset ratio* is an indicator of financial stress, reflecting solvency and operating risks.[7] It is measured by the ratio of total farm and household liabilities to total assets for the year prior the study (1987). Financial hardship should be evident through behavioral adjustments in household consumption. To measure adjustments in consumption expenditures, we used an index developed by Bultena, Lasley and Geller (1986). It was created from affirmative-negative responses to 13 adaptations made by the family in the past year to financial stress such as: postponing major farm purchases; postponing major household purchases; changing food or eating patterns to save money; purchasing more on credit; and postponing medical

care to save money. A simple additive index of "yes" responses was tallied, then divided by the total number of items to calculate an overall score. Cronbach's alpha coefficient for the internal consistency of this measure, *financial adjustments,* is .825. The mean adjustment score for women is .34, indicating that on average, women report that their families made about 34% of the adjustments. We used two indicators of subjective perceptions of financial conditions. *Perceived economic hardship* was measured by a three-item scale from Pearlin et al. (1981), questioning respondents how often in the past year their households could not afford adequate food, clothing, or medical care. Frequent problems in affording household medical care were reported by 11% of women and about 20% reported frequent difficulty in affording needed clothes. Higher scores on the index indicate greater hardship. *Dissatisfaction with finances and quality of life* was composed of five items rating satisfaction with farm family financial and social well-being at present, relative to the past, and for future potentials. Higher scores indicate higher dissatisfaction. Cronbach's alpha coefficients were above .75 for each of the two indicators of subjective perceptions. As we expected from women's roles in mediating family consumption needs, women were significantly more sensitive to deterioration in household conditions than men, reporting greater consumption cutbacks, hardships, and poorer enterprise-household finances (Table 4). We also included family income because it is often used as an independent control variable in models predicting off- and on-farm work and work hours (Buttel and Gillespie 1984; Simpson et al. 1988). Farmers with higher family income were more likely to work off-farm. Net family income refers to annual income from the past year. It was coded into 10 categories to approximate an interval scale.

Sociodemographic Characteristics.

We included other household and personal attributes commonly used in studies of farm and off-farm work that are also germane to the two perspectives on the flexibility and rigidity of women's roles. Human capital characteristics of age, education, and farming experience should influence which household member allocates time to farm or nonfarm employment (Tickamyer and Bokemeier 1988). Age is also important because farmers in their younger stages of farming careers were more adversely affected by the crisis. Level of education was measured by a seven-point scale ranging from none (1) to graduate school (7); over half the sample completed high school (4). Farm background referred to main residence until age 16, (1 nonfarm, 2 farm) and reflected experience with farming as an occupation. About 89% of men and 55% of women had farming backgrounds. Because feminists emphasize property ownership as influencing women's participation in farming (Sachs 1983; Whatmore 1990), we use a scale measuring extent of ownership of capital assets, such as machinery and equipment, which ranged from (1) respondent holds title to none to (4) respondent holds title to all capital assets used on the farm. About 60% of men but only 26% of women had title to all capital assets. A similar scale of land ownership was not used as it had little variation for either men or women. Other control variables were: the presence of a spouse who works off-farm, which increases personal likelihood of off-farm work; and the presence of young children (number of children

aged 6 and younger). No controls for race or ethnicity were included; as is characteristic of the Ohio farm population, nearly all of the sample are white.

Labor Market Context

Farmers' labor allocations may also be affected by geographic context. Proximity to metropolitan areas is often controlled in models of off-farm employment but may not have much influence in Ohio as the state is relatively urban. Metropolitan proximity is a three-point index coded by whether the farm household is situated in (1) a nonmetropolitan county, (2) a nonmetropolitan, metropolitan adjacent county, (3) a metropolitan county. A more feminized nonfarm labor force, measured by the percent of the labor force that is female, may also be related to women's greater participation in off-farm work due to local labor demand. Both variables were measured prior to the crisis period, from the 1980 Census of Population.

Role Stress

Role stress was composed of five items with responses ranging from (1) none/no stress to (4) great feelings of stress. Items asked how frequently the respondent: is concerned with stress; experiences stress "on a day-to-day basis"; feels stress has increased over his/her farming career; has "problems in balancing work and family responsibilities"; experiences "pressure in having too much to do in too little time." The items were initially selected through a principal axis factor analysis for both men and women. Items that had high factor loadings for both men and women were retained and compose the index for which Cronbach's alpha is .82. Women reported somewhat greater stress levels than men.

STATISTICAL MODELS

The results reported in Tables 5-7 are from ordinary least squares regression analyses. Specification of the models followed that of previous studies of farm women's and men's work.[8] We also performed analyses adding other enterprise and spouses' characteristics, including work changes, but these did not alter the relationships presented.[9] Tests for the constancy of parameters (Hardy 1993) as well as the hypotheses to be examined, indicated the need for separate models for men and women.

WOMEN'S AND MEN'S ON-FARM WORK

Few characteristics of the enterprise itself influenced female participation in farm work (Table 5). As expected, women spent more time on dairy farms but other enterprise variables were not significant. The effects of crisis-related household conditions were evident in that women who reported greater financial adjustments or objective cutbacks in household consumption spent more time farming. However, perceptions about farm household conditions showed that women who were more dissatisfied with farm finances and life spent less time in farming ($\beta = -.159$), suggesting increased alienation. Prior farming experience and greater ownership of capital assets were related to greater farm work. Women who were older and those with younger children

TABLE 5: Women's and Men's On-Farm Work

	Farm Work				Changes in Farm Work			
	Women		Men		Women		Men	
	b	B	b	B	b	B	b	B
Farm structure								
Gross farm sales	.007	.001	3.018	.415***	.001	.006	.031	.114*
Hired labor use	-.367	-.013	1.388	.046	.075	.059	-.028	-.025
Cash grains	-2.746	-.070	-3.219	-.077*	.050	.027	-.060	-.038
Dairying	5.455	.161**	1.147	.030	.126	.079	.023	.016
Farm-household financial conditions								
Debts-to-assets	1.777	.037	-2.602	-.048	.066	.031	-.098	-.048
Financial adjustments	12.089	.231***	-3.357	-.054	-.105	-.043	-.190	-.081
Perceived economic hardship	-.304	-.030	.664	.059	.025	.054	.077	.180***
Dissatisfaction with finances	-.633	-.159**	.411	.096*	-.002	-.011	-.007	-.047
Family income	.078	.016	-.476	-.095*	.001	.007	-.005	-.028
Sociodemographic characteristics								
Ownership of capital assets	1.315	.114**	1.620	.092**	.007	.014	.091	.138***
Young children	-3.553	-.177***	.219	.008	.021	.023	.005	.005
Farm background	4.156	.152***	3.414	.079*	.016	.013	.080	.049
Age	-.218	-.205***	-.193	-.182***	-.012	-.258***	-.016	-.401***
Education	-.110	-.008	-.854	-.061	.018	.031	.016	.031
Spouse works off-farm	-2.855	-.104	1.127	.036	-.059	-.046	-.098	-.083
Off-farm work hours	-.174	-.233***	-.183	-.260***	-.030	-.268	NA	NA
Changes in off-farm work	NA	NA	NA	NA	-.297	-.265***	-.217	-.191***
Labor market context								
Metropolitan proximity	.130	.005	.313	.012	.035	.032	.050	.053
Percent labor force female	-13.204	-.075	-7.675	-.035	.031	.003	-.655	-.080
R^2	.210		.412		.137		.205	
Adjusted R^2	.178		.390		.102		.175	

* p < .05 ** p < .01 *** p < .001

reported less time in farming. Women also adjusted their time spent in farming relative to off-farm work, as indicated by the significant negative relationship between off-farm work hours and farm work. For men, larger farms increased time spent in farm work. Men on cash grain farms reported slightly less work time. In contrast to findings for women, greater dissatisfaction with farm household finances and lower family income were significantly related to greater time spent on farm work for men. This suggests a pattern commonly noted in case studies of the crisis period in which the full-time farmer experienced low income and frustration with farm life (Davidson 1990; Heffernan & Heffernan 1986). For women, age, nonfarm background, and greater hours spent in off-farm work reduced time spent farming. Labor market variables had little impact on farm work.

Changes in women's time spent farming over the crisis period were little affected by financial hardship or enterprise structure, supporting the second perspective on the inflexibility of women's farm work. Women who increased their off-farm work time concomitantly decreased farm work time rather than increased total work time in both domains. This conflicted with expectations derived from historical accounts of past crises in which farmers increased total labor time and self-exploitation. Women who owned more capital assets and came from farm backgrounds were no more likely to have increased farm work than other women. In contrast, men's farm work increased as a response to financial hardship and to rationalistic economic criteria. Men who perceived greater economic hardship, whose households operated larger units, and who personally owned greater capital assets were more likely to increase farm work.

In sum, these analyses supported the second perspective of gender rigidity in farm work. Women's labor allocations were only slightly responsive to enterprise exigencies and financial conditions despite a period of downturn. Because men tend to control farm operations, where possible (on large enterprises and for younger farmers), they responded with greater farm work during the crisis. Further, poorer financial conditions associated with the crisis context or feelings of economic hardship and dissatisfaction with finances were linked to greater farm work for men as well. Women remained more distant from farming and did not increase farm work time over the crisis period, even if they owned more assets and came from a farming background.

WOMEN'S AND MEN'S OFF-FARM WORK

Factors related to off-farm work similarly varied by gender with enterprise structure having little relationship to whether or not a women worked off-farm (Table 6). The major determinants of women's off-farm work were sociodemographic: higher education (β = . 240), younger age (β = -.275), the absence of young children (β = -.169), and having a spouse who works off-farm (β = .119). Women reporting greater household consumption adjustments were slightly more likely to work off-farm (β = .102) but this relationship was not quite significant. Coming from a higher income household and from a farm background also increased the propensity for off-farm work. Women who spent more time at farm work were less likely to work off-farm, indicating work domain trade-offs as found in the previous models. As expected, men's off-farm

employment was closely connected to the size of the farm enterprise ($\beta = -.401$). Farm-household financial conditions also influenced men's off-farm work. Men reporting greater household adjustments and higher debt were also more likely to work off-farm. As in the model for women, the presence of a spouse who works off-farm, higher income, and less time spent farming were related to men's off-farm employment. In contrast to women, however, farm background decreased men's off-farm work and education had no significant effect. Probit analyses were also run for this model. As the sample is relatively evenly distributed across the dichotomous dependent variable, the results are the same as for the ordinary least squares (OLS) analysis.

The analysis for time spent per week spent at off-farm work was performed for the 235 women and 200 men who reported off-farm employment. While fewer of the coefficients were significant in part due to smaller sample size, some general patterns were consistent with the previous model. Farm structure again had no significant effects for women, although there was a slight tendency for them to work more hours off-farm on cash grain and smaller units. Demographic variables were similar in direction for women as in the previous model. Women whose households made more objective financial adjustments in the past year worked somewhat longer hours ($p = .08$). Men's work time increased when they had smaller size farms, did not engaged in dairying ($p = .06$), and were slightly more dependent on cash grains. Higher family income and younger age were significantly related to greater hours spent in off-farm work for men. In contrast to findings for women, subjective perceptions of financial dissatisfaction rather than consumer adjustments were associated slightly with longer off-farm work hours for men.

Changes in off-farm work time were analyzed for the full sample. Support for contentions that farm women increased off-farm work due to crisis conditions was seen in the relationships for household-enterprise financial conditions. Women who perceived greater economic hardship and whose households bore higher debt loads reported increased off-farm work. As in previous models, younger age, higher education, and decreased farm work were related to increased time spent in off-farm work for women. Enterprise factors had no effect on women's off-farm work. Labor market variables also show significant relationships for understanding the spatial context within which farm women increased employment. Women more distant from metropolitan areas increased off-farm work, which corresponds to studies noting the greater integration of the most rural women into the labor force during the 1980s (Ollenburger et al. 1989). Women in counties with a more feminized nonfarm labor force also increased off-farm work. Taken together, these relationships suggest that farm women increased labor force participation more in those locales where such trends were set by nonfarm women. For men, changes in off-farm work were related to enterprise exigencies and demographic factors rather than to financial conditions. Men increased their participation in off-farm work on smaller, less commercialized units as indicated by the negative coefficient for hired labor. Younger men also reported increasing off-farm work.

TABLE 6: Women's and Men's Off-Farm Work

	Off-Farm Work				Hours Off-Farm			
	Women		Men		Women		Men	
	b	B	b	B	b	B	b	B
Farm structure								
Gross farm sales	-.007	-.031	-.097	-.401***	-.798	-.113	-3.780	-.344***
Hired labor use	.064	.063	-.035	-.035	-.433	-.015	1.061	.032
Cash grains	-.060	-.042	-.012	-.008	5.570	.126	3.843	.089
Dairying	-.023	-.018	-.075	-.059	-.476	-.013	-7.444	-.128
Farm-household financial conditions								
Debts-to-assets	.114	.065	.147	.082*	.085	.002	-.006	.000
Financial adjustments	.195	.102	.253	.121*	9.237	.180	-8.208	-.122
Perceived economic hardship	.003	.008	-.008	-.023	-1.181	-.116	1.032	.083
Dissatisfaction with finances	.004	.029	-.006	-.043	-.644	-.170	.553	.120
Family income	.024	.144**	.038	.229***	.141	.027	1.700	.256***
Sociodemographic characteristics								
Ownership of capital assets	-.008	-.019	.019	.032	.590	.049	2.557	.135*
Young children	-.124	-.169***	-.014	-.017	-5.017	-.252**	1.006	.440
Farm background	.098	.098*	-.120	-.083*	.701	.025	6.881	.171***
Age	-.010	-.275***	-.012	-.339***	-.167	-.131	-.322	-.233***
Education	.109	.240***	-.005	-.011	1.429	.119	2.028	.127
Spouse works off-farm	.120	.119*	.112	.107**	-3.311	-.117	2.911	.088
Farm work	-.006	-.170***	-.004	-.134**	-.287	-.230**	-.212	-.179*
Changes in farm work	NA	NA	NA	NA	NA	NA	NA	NA
Labor market context								
Metropolitan proximity	-.051	-.060	.021	.025	-1.977	-.083	.365	.013
Percent labor force female	-.058	-.009	-.160	-.022	-8.820	-.046	19.834	.074
R^2	.288		.420		.170		.385	
Adjusted R^2	.260		.400		.100		.325	

* $p < .05$ ** $p < .01$ *** $p < .001$

TABLE 6: Women's and Men's Off-Farm Work (Continued)

	Changes in Off-Farm Work			
	Women		Men	
	b	B	b	B
Farm structure				
Gross farm sales	-.007	-.025	-.018	-.078
Hired labor use	-.054	-.047	-.094	-.095*
Cash grains	.054	.033	-.020	-.015
Dairying	.043	.030	-.021	-.016
Farm-household financial conditions				
Debts-to-assets	.243	.123*	.051	.029
Financial adjustments	-.039	-.018	-.034	-.016
Perceived economic hardship	.070	.169**	.007	.018
Dissatisfaction with finances	-.003	-.022	.004	.034
Family income	.001	.006	.004	.029
Sociodemographic characteristics				
Ownership of capital assets	-.027	-.056	-.021	-.036
Young children	-.058	-.070	-.046	-.057
Farm background	.012	.010	.053	.037
Age	-.010	-.235***	-.014	-.417***
Education	.053	.105**	.015	.033
Spouse works off-farm	-.028	-.024	.015	.015
Farm work	NA	NA	NA	NA
Changes in farm work	-.218	-.244***	-.171	-.196***
Labor market context				
Metropolitan proximity	-.107	-.110**	-.008	-.009
Percent labor force female	.854	.114**	-.263	-.036
R^2	.221		.183	
Adjusted R^2	.180		.152	

* $p < .05$ ** $p < .01$ *** $p < .001$

FARM AND OFF-FARM WORK AND STRESS

To what extent did women's and men's work response during the crisis affect feelings of stress?[10] Women's labor force allocations and enterprise-household conditions were significantly associated with stress: increasing off-farm work, greater financial adjustments, and dissatisfaction with farm household finances and life (Table 7). Poorer household conditions during the crisis coupled with increasing paid employment thus created tension in women's lives. At the same time, paid employment increased women's independence from farming and farm men. Younger, better educated women, the population most at risk during the crisis, also had higher stress. Farm women from larger units and those with higher family incomes, perhaps reflecting the greater need to juggle duties on more affluent enterprises, had slightly higher stress levels. For men, factors related to the farm enterprise were most distinctly related to stress: increasing farm work rather than off-farm work, managing larger units, using more hired labor, and having lower dependency on cash grains, indicating continuous rather than seasonal production of commodities. As with women, financial cutbacks and dissatisfaction increased stress. Stress was also higher among men who reside closer to metropolitan areas, which may be due to the normative environment and complexity of farming near urbanized areas. In sum, this analysis showed that changes in labor allocations were concomitant with social costs, but not only for women as much literature contends.

Conclusions

A large literature has observed that U.S. businesses are responding to changes in the economic environment and conditions of competition by altering traditional divisions of labor and allocating labor more flexibly to production tasks (Piore & Sabel 1984). In contrast to most studies addressing this issue, we focused on family-owned and operated enterprises whose survival has hinged historically on a flexible work force. The 1980s farm sector provided a case in which financial downturn and the agency of household producers would be expected to alter traditional work patterns. We presented two alternative perspectives on women's and men's labor allocations in production during the crisis. One emphasized the allocation of labor based on enterprise features, human capital criteria, and similarities of men's and women's experiences during downturns. It suggested that both women and men should expand productive work with the result that farming may become more feminized. The other emphasized rigidity in family enterprise production and continued or increased marginalization of women in farming. Our findings showed that the way in which farm women and men allocated labor did not correspond to arguments about enterprise flexibility but occurred in a highly rigid way with little change in gendered work domains of farm or household. The consequences of patterns of labor allocations by gender were manifest social-psychologically, in a similarly rigid way.

We first presented an overview of labor allocations, considering the crisis as a context for change. Women tended to increase the amount of time spent in off-farm employment and to decrease the amount of time they spent in most

TABLE 7: Women's and Men's Labor Allocations and Stress

	Role Stress			
	Women		Men	
	b	B	b	B
Farm structure				
Gross farm sales	.174	.100	.212	.122**
Hired labor use	.433	.061	1.352	.189***
Cash grains	-.580	-.057	-1.263	-.126**
Dairying	.474	.054	.065	.007
Farm-household financial conditions				
Debts-to-assets	-.694	.056	-.097	-.007
Financial adjustments	4.298	.317***	3.953	.266***
Perceived economic hardship	.136	.054	.074	.027
Dissatisfaction with finances	.106	.103*	.137	.134**
Family income	.112	.093*	.045	.038
Sociodemographic characteristics				
Ownership of capital assets	.189	.063	.176	.042
Young children	-.022	-.004	.046	.007
Farm background	-.059	-.008	-.155	-.015
Age	-.056	-.204***	-.046	-.182**
Education	.347	.109**	-.055	-.016
Spouse works off-farm	-.116	-.016	.295	.039
Changes in off-farm work	1.070	.172***	.410	.056
Changes in farm work	.131	.024	.749	.118**
Labor market context				
Metropolitan proximity	-.304	-.050	.522	.086*
Percent labor force female	.016	.000	-2.591	-.050
R^2	.361		.342	
Adjusted R^2	.334		.316	

* $p < .05$ ** $p < .01$ * $p < .001$

types of farm work relative to the past and to men. While most Midwest women and men who worked off-farm did so prior to the crisis, over one-third of women and about one-quarter of men reported taking employment due to financial need in the crisis period. In contrast to women's relative marginalization from farming, men were somewhat more likely to expand their farm work roles. Finally, men failed to broaden their scope of activities in household reproduction and despite women's off-farm employment, rarely did household tasks. In sum, there is evidence of gender rigidity and even divergence in farm work, stasis in household work, and convergence in becoming salaried employees.

Within the crisis context, household enterprises occupied varying positions relative to structural characteristics and financial conditions. We questioned to what extent household-enterprise factors are related to patterns of labor allocations. Here, too, factors impelling farm work and semiproletarianization were highly gendered. For women, poorer past enterprise-household conditions were linked to increased off-farm work, thus supporting some analysts' contentions that 1980s expansions in farm women's paid employment were partly crisis-related. For men, poorer enterprise-household conditions were more associated with increased farm work. In contrast to women, men worked longer hours when dissatisfied with farm household finances and life. As other studies found, farm structure had little effect on women's farm and off-farm employment, because men stand at the apex of direct production and management. For men, the size of the enterprise was a consistent determinant of their off-farm work. Men adjusted farm and nonfarm work according to the extent to which farm production could be expanded. For women, sociodemographic attributes were a consistent determinant of their off-farm employment: personal and household factors of higher education, fewer young children, and younger age. Women's farm work was related to sociodemographic attributes as well. However, there was no evidence that women expanded farm work in accordance with rationalistic household or personal factors outlined by perspectives arguing for flexibility in women's farm work. Women worked longer on the farm partly in the face of financial adjustments but also when they perceived improved household finances and quality of life. Women with farming experience failed to increase farm work hours during the crisis, indicating a rigidity of productive roles that belies human capital arguments.

The consequences of family enterprise restructuring also varied differentially by gender. We argued that due to women's position in the family enterprise, they would perceive greater enterprise-household hardship and decline than men, which was confirmed in the Ohio study. Women experienced slightly greater stress levels than men, compounded when their off-farm work, household's financial adjustments, and dissatisfaction with farm finances and life increased. Men's stress increased, in contrast, when time spent in farming rather than nonfarm work increased. Much of the literature on gender and social change has argued that women bear disproportionate costs of economic crises. By demonstrating that patterns of labor allocations, experiences of hardship, and relationship of these factors to stress varied between women and men, this study shows that women bear different rather than disproportionate costs of economic crises.

Despite the absence of wage labor in family based production, the farm sector provides an example where rapid decline has tended to leave the traditional enterprise division of labor intact. Gender differences in labor allocations are reflected even in patterns of stress: for men, increased farm work; for women, increased off-farm work. Why women's marginalization in farming persists is unclear from this study. Our data were limited to the Midwest, with little ethnic variation. The population, though in debt and with lower median income relative to the nation, generally owns most assets and few would be considered impoverished. In the past, women of color and poorer women have been more involved in direct production. Some researchers indicate that the institutional legacy of the agricultural system, government agencies and programs, farming organizations and cooperatives, and agribusiness firms, have been barriers to women's involvement, if not in the present, then certainly in the past (Rosenfeld 1985; Sachs 1983). Others have looked more to the dynamics of the family and to the extent of patriarchal control, values, and norms therein (Barlett 1993; Salamon 1992; Whatmore 1990). Women's labor allocations can also be seen as exiting from an organization of work that men have dominated historically. Women who work off-farm contribute to a fair proportion of household income, which in other contexts, has been shown to increase their power in the family. Had they increased farm work, instead, their labor time and output would have been more closely under the control of farm men.

Conceptually, the findings have implications for the linkage between human agency and social structure. Family enterprises, particularly in farming, have been viewed as holistic economic actors responding to economic change. Scholars have treated the "petit bourgeoisie" as "something like a 'family class'," whose independence and livelihood depends on the voluntary exploitation of family members and on the shared statuses that productive labor affords (Stock 1992:14). Women's unpaid labor is assumed to be highly elastic and readily available to compensate for labor needs and consumption shortfalls. The insight provided by our study is that prior social structural positions, as indicated by gender, are embedded within the family as an economic agent. The type of response to macroeconomic conditions is contingent on the gendered nature of positions. In brief, gender guides the ways in which macrolevel crises are experienced and ways in which responses are manifest.

The findings also have implications for the reflexive nature of agency and social structure. Not only does structure affect actors' lives but also their response alters this structure. Farm women's labor force allocations and household experiences have implications for the future trajectory of family-based production. The strategies that have enabled family farmers to survive and to compete with large, industrialized farms appear to be weakening. During the crisis, farm women and men did not increase both farm and off-farm work but traded time spent in one arena for that in the other. Farm work has been exchanged for nonfarm employment with women more likely to make this adjustment. Another typical strategy, reducing household consumption, is related to higher stress. The experiences of coping with fluctuating financial conditions and other aspects of farm production make social-psychological depression high among the farm population relative to nonfarm people and puts farming in the top ranks of high stress occupations. Off-farm employment

has long been viewed as a viable and relatively stable strategy that could maintain the enterprise in the face of economic downturns and support the farming lifestyle. Economic projections, however, are for the highest future losses in farming to come from smaller enterprises more dependent on off-farm income (Stanton & Olson 1990). Women's increased off-farm employment, somewhat higher stress, greater perceptions of financial hardship and farming dissatisfaction, suggest that farm women will be less likely to continue to underwrite the costs of direct food production relative to their male partners and past history.

Notes

1. The literatures reviewed in this section are framed in terms of both position and period effects. Analysts studying labor allocations by gender during periods of decline generally do not make explicit their conceptual strategy. Three analytic strategies can be seen. The most common strategy is to focus on a particular historical period and to assume that economic changes wash broadly over populations. Examples are studies on: labor allocations during the Depression (Elder & Liker 1982) and previous farming crises and downturns (Friedmann 1978; Stock 1992); women's employment survival strategies in Third World societies undergoing structural adjustment (Moser 1992); and work patterns during recession (Cummings 1987). With the exception of some historical work, these studies are based on cross-sectional samples and assume that the present economic environment is the structural context for studying the manifestations of macro-level changes. A second strategy (which some studies above utilize as well) is to focus on actors' positions which are assumed to result from economic decline, such as gradations of hardship and class variations. Cummings (1987) for example, examines women's labor allocations in industries assumed to have undergone different rates of restructuring, maintaining that the cross-sectional study provides a post-test of the effects of recession. Because nearly all studies on the farm crisis are cross-sectional, they assume that its effects are manifest over the industry at large during this period and also in the gradations of economic positions in which farm households are situated (e.g., the volume by Clay & Schwarzweller 1991). Finally, a third strategy used by a few ethnographic researchers centers on case studies of household labor allocations during the crisis (Barlett 1993; Salamon 1992).

2. Neoclassical economists suggest a number of ways by which economic downturn may be related to greater labor allocations of household members to productive work: objective declines in household income; objective differences between household income and the average income in the area of residence; perceived current income inadequacy or inadequacy relative to the past; and perceived income-consumption gap between one's household and reference group (Rubin & Riney 1994). Perceptions of income adequacy are further assumed to be shared among household members. In sum, although there is no single mechanism by which economic downturn is seen to affect household labor allocations, there is agreement over the end result of this process.

3. A summary of studies that examine stress during the farm crisis can be found in Wagenfeld et al. (1993). Most focus on farm-household economic position and stress and make similar assumptions as research on non-farm populations, that those in poorer positions during downturns have higher stress (Aneshensel 1992). The focus of this paper is not to specify the determinants of stress but to extend the literature on gendered divisions of labor by examining their extra-economic consequences.

4. The population list from which the North Central sample is drawn was provided by the state Agricultural Statistics Services. It is the most comprehensive list of farms available and covers the farms that are in the Census of Agriculture. In order to use the population list, we had to solicit independent cooperation from each of the 12 state Agricultural Statistics Service directors. Due to new confidentiality restrictions imposed by the Service, we were not allowed direct access to the list; we complied with the Statistical Service's request that they mail out the questionnaires and return them to us with no information identifying the respondents' name

or address. The Service directors further indicated that: there would be one mailing of both questionnaires and one reminder postcard; and that the Services would complete telephone follow-up interviews of a random sample of nonrespondents (approximately 50) in each state to examine the issue of potential nonresponse bias. The response rate for the survey was estimated at 45%. To examine the issue of potential nonresponse bias, items from the survey and telephone questionnaires were compared to determine if there were differences in response patterns. No significant differences were found between the two samples with regard to major background variables such as farm size, debt levels, household demographic characteristics and to attitudinal questionnaire items. The two samples (mail and telephone) were then combined for each state and a regional data set, consisting of 4,087 production questionnaires and 3,630 partner questionnaire, was created.

5. About 10% of Midwest and 8% of the Ohio sample are aged 70 or older. We included these respondents in Tables 1,2, and 3 because: each sample is based on the presence of a respondent who defines him/herself as currently operating a farm for the study year; and some farm people over age 70 have off-farm jobs. The figures reported in the table differ little (1-2%) when respondents aged 70 or over and respondents intending to retire from farming are excluded.

6. Studies that link historical context to households and individuals typically cannot do so directly and must infer that individuals' hardships experienced were related to the macroeconomic environment. Such studies "emphasize the ways that larger social structural forces constrain, and to some extent determine, the repertoire of adaptations available to individual families in a given society. These social structural forces have impact not only on the adaptations that are possible, but also on which families — and which individuals within families — receive the most benefit from a given strategy" (Moen & Wethington 1992:243). Macroeconomic contexts thus are seen as an opportunity structure, circumscribing as well as differentially impacting individuals and households. In the present study, we use farm-household conditions to indicate varying positions within farm households that form the basis for members' labor allocations during the crisis. While we recognize that due to the nature of survey research, these conditions cannot be linked directly to the macro-level crisis, we believe that they do at least implicitly capture crisis effects for the following reasons. First, the purpose of both the Ohio and Midwest surveys were stated to the respondents as being conducted to assess the impact of the farm crisis. Second, for the Ohio survey, we specifically phrased all questions measuring household adjustments, dissatisfaction with finances, and perceptions of household hardship — in terms of "changes made due to financial strains" in recent, past years. Items measuring financial conditions — debt-to-asset and income also refer to the past year (1987) while the dependent variables are 1988 data. We thus use retrospective data to tap crisis experiences. While sociologists acknowledge that retrospective data can be problematic due to personal memory, the temporal period was the recent past (generally one to three years), so that this should minimize such issues. Third, unlike many studies that examine period effects only, through examining positions as well, we are able to control for factors extraneously related to the crisis, that could have affected "enterprise-household financial conditions," such as age number of children, and all farm structure variables.

7. The debt-to-asset ratio is a commonly used indicator of financial solvency and reflects the risk associated with the operations' financial structure. However, analysts caution that it is only one measure of a farm's financial conditions: it does not take into account how debt is structured which influences a borrower's ability to service and repay; debt service commitments may be met with earnings which vary by household; and as noted, in the case of the Ohio sample, the debt-to-asset ratio is associated with farm size, with larger farm more highly leveraged but also generally more able to service debt. The debt-to-asset ratio thus should be used in combination with other income and farm size measures, as we have in the regression analyses. The mean debt-to-asset ratio for 1987 was 19% as compared to 15% for the nation (Morehart et al. 1988:134). There is no indication that Ohio farmers became financially stressed in 1987 due to land purchases as we examined the zero-order correlation between the debt-to-asset ratio and a question about purchases of land in the recent past. Our data also indicate that the debt-to-asset ratio taps financial stress for our population. Farmers with debt-to-asset ratios over 40% were significantly more likely to have: increased farm and off-farm work; greater dissatisfaction with farm finances; made more household adjustments; experienced

greater economic hardship; and higher role stress. Those with higher debt-to-asset ratios are also more likely to work off-farm and to have a spouse who does so.

8. We follow Simpson et al. (1988) and Buttel and Gillespie (1984) by including time spent in farm work as an independent variable in the models for farm work; and by including time spent in off-farm work as a determinant of the amount of time spent in farm work. Reasons for this noted by the above studies are: to control for labor time spent in another work domain; to test for gender differences in these relationships. Including these variables is also relevant due to patterns of labor allocation observed during the farm crisis: farm women and men were said to increase total work hours, both on the farm and off-farm, which, in turn, compounded personal and family stress. Since our analysis collects only cross-sectional data on the independent variables, we are unable to ascertain causality. Rather, our focus is limited to delineating characteristics that are related to patterns of work and changes in work during the crisis period. However, it should be noted that most independent variables tapping household enterprise financial conditions and enterprise and demographic characteristics were framed to the respondent as causally prior to the dependent variables or could be considered causally prior, such as capital ownership.

9. We performed additional analyses to include variables on the extent of landownership and the availability of other family labor. Adding these variables to the regression models in Tables 5, 6, and 7 resulted in no significant relationship with any of the dependent variables nor did these variables alter any of other relationships in these models. We also performed analyses adding variables about the respondents' spouse. In this case, the relevant analyses would be for the 86% of the sample that is couples only. We created new variables for men and women adding their spouses' perceptions of economic hardship, amount of farm labor performed, and changes in farm and off-farm work. While spouses' variables have some influence on respondents' labor allocations and stress, all relationships reported in Tables 5, 6, 7, stay the same when the above variables are added and the sample is confined to only couples.

10. Specification of the models for stress follows analyses presented in previous tables. This allows us to control for all variables that might affect changes in labor allocations, in turn, permitting the direct effects of these variables on stress to be measured. The contextual variables are kept in this last analysis for reasons of continuity. The types of variables in the model for stress are those that have been used previously in studies of nonfarm populations (e.g., economic hardship and demographic variables) and have been used as well in studies of farmers during the crisis. Contextual variables are not often included in models of stress, although there is increased recognition that they should be (Aneshensel 1992; Wagenfeld et al. 1993). In brief, local context has been seen as providing economic resources, social support, and a normative environment which affects stress. In the sociological literature, a longstanding conceptual argument has been that areas which are more urbanized or less rural offer a normative environment associated with additional stressors. Similarly, areas where a higher proportion of women are in the labor force may be related to lower personal stress if one considers this an indicator of the normative work environment, economic resources, and opportunities.

References

Aneshensel, Carol S. 1992. "Social Stress: Theory and Research." *Annual Review of Sociology* 18: 15-38.

Aronson, Robert L. 1991. *Self-Employment: A Labor Market Perspective*. ILR Press.

Bagguley, Paul, Jane Mark-Lawson, Dan Shapiro, John Urry, Sylvia Walby, Alan Warde. 1990. *Restructuring: Place, Class, and Gender*. Sage.

Barlett, Peggy. F. 1993. *American Dreams, Rural Realities: Family Farms in Crisis*. University of North Carolina Press.

Becker, Gary S. 1981. *A Treatise on the Family*. Harvard University Press.

Belyea, Michael J. and Linda M. Lobao. 1990. "Psychosocial Consequences of Agricultural Transformation: The Farm Crisis and Depression." *Rural Sociology* 55:58-75.

Bultena, Gordon, Paul Lasley, and Jack Geller. 1986. "The Farm Crisis: Patterns and Impacts of Financial Distress Among Iowa Farm Families." *Rural Sociology* 41:391-405.

Buttel, Frederick H. and Gilbert W. Gillespie. 1984. "The Sexual Division of Farm Household Labor: An Exploratory Study of the Structure of On-Farm and Off-Farm Labor Allocation among Farm Men and Women." *Rural Sociology* 49:182-209.

Castells, Manual, and Alejandro Portes. 1989. "World Underneath: The Origins, Dynamics, and Effects of the Informal Economy." Pp. 11-37 in *The Informal Economy: Studies in Advanced and Less Developed Countries*, edited by Alejandro Portes, Manuel Castells, and Lauren A. Benton. Johns Hopkins University Press.

Clay, Daniel, and Harry K. Schwarzweller (eds.). 1991. *Research in Rural Sociology and Development: Household Strategies*. JAI Press.

Collins, Jane L. 1990. "Unwaged Work in Comparative Perspective: Recent Theories and Unanswered Questions." Pp. 3-24 in *Work Without Wages: Comparative Studies of Domestic Labor and Self-Employment*, edited by Jane L. Collins and Martha Gimenez. SUNY Press.

Conger, Rand D. and Glen H. Elder Jr. 1994. *Families in Troubled Times: Adapting to Change in Rural America*. Aldine de Gruyter.

Coughenour, C. Milton and Louis E. Swanson. 1983. "Working Statuses and Occupations of Men and Women in Farm Families and the Structure of Farms." *Rural Sociology* 48:23-43.

Coverman, Shelley. 1989. "Role Overload, Role Conflict, and Stress: Addressing Consequences of Multiple Role Demands." *Social Forces* 67:965-82.

Cummings, Scott. 1987. "Vulnerability to the Effects of Recession: Minority and Female Workers." *Social Forces* 65:834-57.

Daines, Victoria and David Seddon. 1993. "Confronting Austerity: Women's Responses to Economic Reform." Pp. 3-32 in *Women's Lives and Public Policy: The International Experience*, edited by Meredeth Turshen and Briavel Holcomb. Greenwood Press.

Davidson, Osha Gray. 1990. *Broken Heartland*. Anchor Books.

de Janvry, Alain, and Phillip LeVeen. 1986. "Historical Forces that Have Shaped World Agriculture: A Structural Perspective." Pp. 83-104 in *New Directions for Agriculture and Agricultural Research: Neglected Dimensions and Emerging Alternatives*, edited by Kenneth A. Dahlberg. Rowman & Allanheld.

Duncan, S.F., R.J. Volk, and R.A. Lewis. 1988. "The Influence of Financial Stressors upon Husbands' and Wives' Well-being and Family Life Satisfaction," Pp. 32-39 in *Families in Rural America: Stress, Adaptation and Revitalization*, edited by R. Martoz-Baden, C.B. Hennon, and T.H. Brubaker. St. Paul, Minn.: National Council on Family Relations.

Edwards, Richard. 1979. *Contested Terrain: The Transformation of the Workplace in the Twentieth Century*. Basic Books.

Elder, Glen H., Jr., and J. Kurt Liker. 1982. "Hard Times in Women's Lives: Historical Influences across 40 years." *American Journal of Sociology* 88:241-69.

England, Paula, and Barbara Stanek Kilbourne. 1990. "Markets, Marriage, and Other Mates: The Problem of Power." Pp. 163-88 in *Beyond the Marketplace: Rethinking Economy and Society*, edited by Roger Friedland and A.F. Robertson. Aldine de Gruyter.

Findeis, Jill L., Daniel A. Lass, and Milton C. Hallberg. 1991. "Effects of Location on Off-farm Employment Decisions." Pp. 263-76 in *Multiple Job-holding among Farm Families*, edited by Milton C. Hallberg, Jill L. Findeis, and Daniel Lass. Iowa State University Press.

Friedland, William H. 1991. "Women and Agriculture in the United States: A State of the Art Assessment." Pp. 315-38 in *Towards a New Political Economy of Agriculture*, edited by William H. Friedland, Lawrence Busch, Frederick Buttel, and Allan Rudy. Westview.

Friedmann, Harriet. 1978. "World Market, State, and Family Farm: Social Bases of Household Production in an Era of Wage Labor." *Comparative Studies in Society and History* 20:545-86.

Gladwin, Christina. 1991. "Multiple Job-holding among Farm Families and the Increase in Women's Farming." Pp. 213-35 in *Multiple Job-holding among Farm Families*, edited by M.C. Hallberg, Jill L. Findeis, and Daniel Lass. Iowa State University Press.

Hanson, Gregory. 1990. "Beyond the Farm Debt Crisis." *Choices* 4th qtr.:33-35.

Hardy, Melissa A. 1993. *Regression with Dummy Variables*. Sage.

Hareven, Tamara. 1990. "A Complex Relationship: Family Strategies and the Processes of Economic and Social Change." Pp. 215-244 in *Beyond the Marketplace: Rethinking Economy and Society*, edited by Roger Friedland and A.F. Robertson. Aldine de Gruyter.

Harl, Neil E. 1990. *The Farm Debt Crisis of the 1980s*. Iowa State University Press.

Heffernan, William D. and Judith B. Heffernan. 1986. "The Farm Crisis and the Rural Community." Pp. 273-80 in *New Dimensions in Rural Policy: Building Upon Our Heritage*, edited by Dale Jahr, Jerry W. Johnson, and Ronald C. Wimberley. Studies prepared for the use of the Subcommittee on Agriculture and Transportation of the Joint Economic Committee, Congress of the United States. Government Printing Office.

Hochschild, Arlie. 1989. *The Second Shift: Working Parents and the Revolution at Home*. Viking.

Huffman, Wallace E. 1991. "Agricultural Household Models: Survey and Critique." Pp. 79-111 in *Multiple Job-holding among Farm Families*, edited by M.C. Hallberg, Jill L. Findeis, and Daniel Lass. Iowa State University Press.

Johnson, Gloria, Paul Lasley, and Kevin A. Kettner. 1991. "Hardship and Adjustment Among Farm Households in Iowa." Pp. 211-22 in *Research in Rural Sociology and Development Volume 5: Household Survival Strategies*, edited by Daniel C. Clay and Harry Schwarzweller. JAI Press.

Jones, JoAnn, and Rachel A. Rosenfeld. 1989. "Women's Occupations and local Labor Markets: 1950-1980. *Social Forces* 67:666-92.

Jones, John, and Patrick Canning. 1993. *Farm Real Estate Historical Series Data, 1950-1992*. USDA, ERS (Statistical Bulletin no. 855), May.

Kenney, Martin, Linda M. Lobao, James Curry, and W. Richard Goe. 1989. "Midwestern Agriculture in U.S. Fordism: From New Deal to Economic Restructuring." *Sociologia Ruralis* 29:130-48.

Krause, Kenneth. 1987. *Corporate Farming, 1969-1982*. Agricultural Economic Report no. 578. Washington, D.C.: USDA, ERS.

Lasley, Paul, F. Larry Leistritz, Linda Lobao, and Katherine Meyer. 1995. *Beyond the Amber Waves of Grain: An Examination of Social and Economic Restructuring in the Heartland*. Westview.

Leistritz, F. Larry and Steve H. Murdock. 1988. "Financial Characteristics of Farms and of Farm Financial Markets and Policies in the United States." Pp. 13-28 in *The Farm Financial Crisis: Socioeconomic Dimensions and Implications for Producers and Rural Areas*, edited by S.H. Murdock and F.L. Leistritz. Westview Press.

Lipman-Blumen, Jean. 1984. *Gender Roles and Power*. Prentice-Hall.

Lobao, Linda M. 1990. *Locality and Inequality: Farm and Industry Structure and Socioeconomic Conditions*. SUNY Press.

Lyson, Thomas A. 1985. "Husband and Wife Work Roles and the Organization and Operation of Family Farms." *Journal of Marriage and the Family* 47:759-64.

McDonald, Heather and Alan Peters. 1991. "Farm Women in the New Rural Labor Markets of the West North Central Region." Pp. 51-73 in *Proceedings of the 1991 Regional Conference on The Rural Family, the Rural Community, and Economic Restructuring*, edited by Ken Root, Judy Heffernan, Gene Summers, and Julie Stewart. Ames, Iowa: The North Central Regional Center for Rural Development.

Meyer, Katherine and Linda Lobao. 1994. "Engendering the Farm Crisis: Women's Political Response in the USA." Pp. 69-86 in *Gender and Rurality*, edited by Sarah Whatmore, Terry Marsden and Philip Lowe. David Fulton.

Mingione, Enzo. 1991. *Fragmented Societies: A Sociology of Economic Life beyond the Market Paradigm*. Basil Blackwell.

Moen, Phyllis and Elaine Wethington. 1992. "The Concept of Family Adaptive Strategies." *Annual Review of Sociology* 18:233-51.

Mooney, Patrick. 1988. *My Own Boss? Class, Rationality and The Family Farm*. Boulder: Westview.

Morehart, Mitchell, James D. Johnson, and David E. Banker. 1988. *Financial Characteristics of U.S. Farms, January 1, 1983*. Washington, D.C.: Economic Research Service, USDA Agriculture Information Bulletin no. 551.

Moser, Caroline. 1992. "Adjustment from Below: Low-Income Women, Time and the Triple Role in Guayaquil, Ecuador." Pp. 87-116 in *Women and Adjustment Policies in the Third World*, edited by Haleh Afshar and Carolyne Dennis. St. Martin's Press.

Murdock, Steve H., and F. Larry Leistritz. 1988. " Introduction." Pp. 1-9 in *The Farm Financial Crisis: Socioeconomic Dimensions and Implications for Producers and Rural Areas*, edited by Steve H. Murdock and F. Larry Leistritz. Westview Press.

Ollenburger, Jane C., Sheryl J. Grana, and Helen A. Moore. 1989. "Labor Force Participation of Rural Farm, Rural Nonfarm, and Urban Women." *Rural Sociology* 54:533-50.

Pearlin, Leonard I., Morton A. Lieberman, Elizabeth G. Menaghan, and Joseph T. Mullan. 1981. "The Stress Process." *Journal of Health and Social Behavior* 22:337-56.

Pfeffer, Max J. 1989 "The Feminization of Production on Part-time Farms in the Federal Republic of Germany." *Rural Sociology* 54:60-73.

Piore, Michael, and Charles Sabel. 1984. *The Second Industrial Divide*. Basic Books.

Pollak, Robert, and M.L. Watcher. 1975. "The Relevance of the Household Production Function and Its Implications for the Allocation of Time." *Journal of Political Economy* 83:255-78.

Ritchie, Martha Farnsworth, and Kelvin M. Pollard. 1994. "U.S. Women Are Anything but 'Typical'." *Population Today* 22:4-5.

Rosenblatt, Paul C. 1990. *Farming Is in Our Blood: Farm Families in Economic Crisis*. Iowa State University Press.

Rosenfeld, Rachel A. 1985. *Farm Women: Work, Farm, and Family in the United States*. University of North Carolina Press.

Rubin, Rose M., and Bobye J. Riney. 1994. *Working Wives and Dual-Earner Families*. Praeger.

Sachs, Carolyn E. 1983. *The Invisible Farmers: Women in Agricultural Production*. Rowman & Allanheld.

Salamon, Sonya. 1992. *Prairie Patrimony: Family, Farming, and Community in the Midwest*. The University of North Carolina Press.

Shaver, Frances M., and Bill Reimer. 1991. "Economy and Household: The Gender-Based Division of Labor on the Farm." Pp. 131-47 in *Research in Rural Sociology and Development: A Research Annual*. Vol. 5, edited by Harry K. Schwarzweller and Daniel Clay. JAI Press.

Simpson, Ida Harper, John Wilson, and Kristina Young. 1988. "The Sexual Division of Farm Household Labor: A Replication and Extension." *Rural Sociology* 53:145-65.

Spitze, Glenna. 1986. "The Division of Task Responsibility in U.S. Households: Longitudinal Adjustments to Change." *Social Forces* 64:689-701.

Stam, Jerome M., Steven R. Koenig, Susan E. Bentley, H. Frederick Gale Jr. 1991. *Farm Financial Stress, Farm Exits, and Public Sector Assistance to the Farm Sector in the 1980s*. Agricultural Economic Report no. 645, Washington, D.C.: USDA, ERS.

Stanton, B.F., and K.D. Olson. 1990. "The Impacts of Structural Change and the Future of American Agriculture." Pp. 261-90 in *Determinants of Farm Size and Structure*, edited by Arne Hallam. Iowa State University Press.

Stock, Catherine McNicol. 1992. *Main Street in Crisis: The Great Depression and the Old Middle Class on the Northern Plains*. University of North Carolina Press.

Tickamyer, Ann, and Janet Bokemeier. 1988. "Sex Differences in Labor Market Experiences." *Rural Sociology* 53:166-89.

Tilly, Louise A. 1981. "Paths of Proletarianization: Organization of Production, Sexual Division of Labor, and Women's Collective Action." *Signs* 7:400-17.

U.S. Bureau of the Census. 1991. *Statistical Abstract of the United States: 1991*. 11th ed. Washington, D.C.

Wagenfeld, Morton O., J. Dennis Murray, Dennis Mohatt, and Jeanne C. DeBruyn. 1993. *Mental Health in Rural America: 1980-1993*. National Institutes of Mental Health.

Walby, S., and P. Bagguley. 1989. "Gender Restructuring: Five Labor Markets Compared." *Environment and Planning D: Society and Space* 7:277-92.

Walker, Lilly Schubert, and James L. Walker. 1988. "Stressors and Symptoms Predictive of Distress in Farmers." *Family Relations* 36:374-78.

Whatmore, Sarah. 1991. *Farming Women: Gender, Work, and Family Enterprise*. Macmillan.

Wilson, John, Ida Harper Simpson, and Richard Landerman. 1994. "Status Variation on Family Farms: Effects of Crop, Machinery, and Off-farm Work." *Rural Sociology* 59:136-53.

Attitudes, Values, and Entrance into Cohabitational versus Marital Unions*

MARIN CLARKBERG, *University of Chicago*
ROSS M. STOLZENBERG, *University of Chicago*
LINDA J. WAITE, *University of Chicago*

Abstract

This article argues that marriage and cohabitation are associated with important differences in work patterns, earnings, treatment of money, use of leisure time, social relations with the extended family, the division of household labor, and fertility. We hypothesize that these differences lead those considering the formation of a household to consider their attitudes toward these aspects of life, which appear to be so different in marriage from those in cohabitation. Using data from the National Longitudinal Study of the High School Class of 1972, we test and find support for the hypothesis that the choice between cohabitation and marriage is affected by attitudes and values toward work, family, use of leisure time, money, and sex roles, as well as values and attitudes toward marriage itself.

Cohabitation is now an important form of union formation in the U.S. Recent estimates suggest that about 35% of those born in the early 1960s will live with someone of the opposite sex before age 25 compared to less than 8% of those born in the early 1940s (Bumpass & Sweet 1989). Although cohabitation is now unexceptional, key questions remain about why some who form households establish cohabitational unions while others marry. This article argues that the choice between marriage and cohabitation impinges on a wide range of activities, circumstances, and social relations. These most obviously include procreation and relations between cohabitants or spouses, but we argue that they also comprise relations with extended family, leisure pursuits, employment, division of economic resources, and division of household labor. Cohabiting may allow individuals or couples who feel unready for the demands of marriage to delay the assumption of marital roles but to acquire the benefits

** The research reported here was supported by Center Grant #P-50-HD-12639 from the Center for Population Research, National Institute of Child Health and Human Development to the Population Research Center, the RAND Corporation. We would like to thank Richard Baumann and the RAND Population Research Center Data Core for preparation of the data set. Direct correspondence to Marin Clarkberg, Population Research Center, 1155 E. 60th Street, Chicago, IL 60637. E-mail: aaprini@mazel.spc.uchicago.edu.*

of coresidence. Consequently, we hypothesize that the choice between cohabitation and marriage for first union is affected by attitudes and values toward work, family, use of leisure time, money, and sex roles, as well as values and attitudes toward marriage itself. To explore this general hypothesis, we apply discrete-time event-history methods to data on the marriage and cohabitation patterns of male and female respondents to the National Longitudinal Study of the High School Class of 1972.

Background: Cohabitation versus Marriage

Unmarried cohabitation (hereafter "cohabitation") in the past was often called "common law marriage" and often has been treated as simply a briefer (Willis & Michael 1994) or less legalistic (Teachman & Polonko 1990; Teachman, Thomas & Paasch 1991) variety of marriage. However, recent longitudinal evidence indicates that values and attitudes toward marriage affect the propensity of individuals to form cohabitational unions (Axinn & Thornton 1992). Other studies (discussed below) report substantial differences between the individual characteristics, time-use patterns, and values and attitudes of married and cohabiting persons. Interpersonal dynamics of marriages and cohabitational unions also have been found to differ. However, some of the most suggestive of these analyses are based upon cross-sectional data that cannot distinguish the effects of attitudes and values on cohabitation from the effect of cohabitation on values and attitudes (e.g., Bumpass, Sweet & Cherlin 1991; Lesthaeghe & Surkyn 1988; Rindfuss & VandenHeuvel 1990; Thomson & Colella 1992). Some key findings follow.

TIME USE

Economic and sociological theories often conceive of marriage as an institution in which the pursuit of individual goals is replaced by joint utility-maximization by husband and wife (e.g., Becker 1991; Parsons & Bales 1955). Some of these goals are merely joint versions of individual goals, such as obtaining food, shelter, and other necessities. Other goals can exist only in the context of couple relationship, such as childbearing (Rindfuss & VandenHeuvel 1990) and fidelity (Caplow, Bahr & Chadwick 1983) or are correlated with marriage, e.g., regular religious attendance and home ownership (Caplow, Bahr & Chadwick 1983; Rindfuss & VandenHeuvel 1990). In theory, replacement of individual objectives by joint goals changes the time allocation patterns of actors; empirical studies (cited below) find that it does so.

Many time-use group studies amalgamate cohabitors with single persons and therefore can only suggest differences between cohabitors and married persons. But this literature does indicate that time spent on work, housework, and leisure all vary substantially by marital status (Juster & Stafford 1985). Most notable is time spent in the labor market: married men spend significantly more time in the labor market than single men and married women spend significantly less than their single counterparts (e.g., Juster & Stafford 1985; Shelton 1992). Women's time in the labor market drops as their time spent in housework increases: even among employed women, marriage increases time spent on

household labor by a third or more (Shelton 1992; South & Spitze 1994). Even in two-profession couples where both partners devote similar numbers of hours to paid labor, wives are still far more likely to change or leave jobs due to a geographic change in husbands' place of work than vice-versa (Bielby & Bielby 1992). And finally, while single and married women spend comparable amounts of time on leisure activities, men experience large drops in leisure time, particularly active leisure, with marriage (Juster & Stafford 1985; Shelton 1992).

Empirical studies find that cohabitors segregate household tasks less by gender than do married people (Blair & Lichter 1991). And while female cohabitors still spend more time on housework than male cohabitors, male cohabitors spend more time on housework than either single men or married men, and cohabiting women spend less time at housework than married women (South & Spitze 1994). To the extent that husbands do contribute to housework, their efforts are often considered "helping out" and their wives typically continue to have primary responsibility to ensure that the task is carried out (Berk 1985; Hochschild 1989). In addition, cohabitors spend less time on religious activities than married persons (Stolzenberg, Blair-Loy & Waite 1995; Thornton, Axinn & Hill 1992). And Blumstein and Schwartz (1983) note that "Cohabitors feel most strongly about having time away from their partners . . . [and] have the clearest sense of a need for personal time. . . . Their belief in personal independence . . . encourages them to maintain separate lives" (186).

INTERPERSONAL DYNAMICS

Focusing more on cohabitations and marriages than on the individuals who cohabit and marry, Brines and Joyner (1993) and Blumstein and Schwartz (1983) argue that cohabitational unions are bound together by egalitarian individualism instead of the joint utility maximization that characterizes marriage in its widely accepted models (Becker 1991; see also Parsons & Bales 1955). Blumstein and Schwartz (1983) find that cohabitors tend to explicitly agree that both partners should contribute equally to all areas of their relationship. Brines and Joyner (1993) find that income inequalities between partners tend to destabilize cohabitations while doing the opposite to marriages. This finding led Brines and Joyner to conclude that "marriage and cohabitation differ not merely by legal status or by the attributes of partners, but also by the internal processes that promote cohesion" (35).

INDIVIDUAL CHARACTERISTICS

Rindfuss and VandenHeuvel (1990) argue that cohabitors are more likely to describe themselves as "never married" than "married" and to have nonfamilial activity patterns and home ownership rates more characteristic of unmarried noncohabitors than of married persons. Fertility patterns of cohabitors differ substantially from those of the married (Bachrach 1987). Cohabitors differ from married persons in demographic characteristics such as race, ethnicity, income, and education (Bumpass & Sweet 1989; Glick & Spanier 1980); illicit drug use (Yamaguchi & Kandel 1985); employment experience (Clarkberg 1994); and even "personality problems" (Booth & Johnson 1988).

VALUES AND ATTITUDES

Lesthaeghe and Surkyn (1988) and Lesthaeghe, Moors, and Halman (1993) find that cohabitors are less likely than married persons to disapprove of marital infidelity and divorce. Cohabitors are less inclined than married persons to espouse traditional gender roles (Lesthaeghe & Surkyn 1988) and more likely to stress the importance of individual freedom within marriage (Thomson & Colella 1992). Compared to married persons, cohabitors expect fewer children (Rindfuss & VandenHeuvel 1990), express less commitment to an ideal of marital permanence (Axinn & Thornton 1992; Bumpass, Sweet & Cherlin 1991; DeMaris & MacDonald 1993), and are less attached to the principle of marital sexual fidelity (DeMaris & MacDonald 1993). Studies by Axinn and Thornton (1992) and Thornton, Axinn, and Hill (1992) are particularly noteworthy because they use longitudinal data to examine the effects of religiosity and attitudes toward divorce on the propensity to subsequently cohabit or marry. These analyses find that a person's religiosity and disapproval of divorce have positive effects on the probability that his or her first union is a marriage rather than an unmarried cohabitation.

Hypotheses: Attitudes and Values

Although little of the previous research on the determinants of cohabitation supports strict causal interpretation, the literature reviewed above suggests that cohabitation and marriage are associated with large differences in work patterns, earnings, treatment of money, time use, leisure, social relations with extended family, the division of household labor, fertility, and perhaps other things too. The well-defined and often sex-typed roles and responsibilities of ideal marriage contrast sharply with the emphasis on equality, independence, self-reliance, and self-determination that seems to characterize cohabitation. These differences lead us to believe that rational people contemplating formation of a household would consider their attitudes toward the aspects of life that appear to be so different in marriages from those in cohabitations. Other things equal, these people would choose to marry or cohabit, depending upon the extent to which marriage or cohabitation appears more likely to produce their current preferred patterns of time use, contact with relatives, work, leisure, and so on.

Cohabitation tends to be a relatively short-lived union. Most cohabiting couples either marry or separate within only two years (Bumpass & Sweet 1989). But some who wish to marry eventually may prefer to delay the assumption of marital roles, we argue, precisely because they value leisure time, money, career success, or other goals not compatible with marriage in the short run. Thus, some individuals may, because of their attitudes and values, enter a cohabitation as a step toward marriage, but one that offers them more freedom in the short term than marriage would. Others may prefer never to assume the breadwinner or housewife role and may prefer long-term cohabitation for this reason. But in both cases, current attitudes and values toward time use, interpersonal relationships, career, and money affect the chances of forming a

union and the choice to marry directly or cohabit, perhaps in anticipation of later marriage.

Thus, we hypothesize that attitudes and values toward work, family, use of leisure time, money, and sex roles — as well as values and attitudes toward marriage itself — all affect the probability that an individual who forms a household for the first time will do so as a cohabitation rather than a marriage.[1]

IMPORTANCE OF MARRIAGE, FAMILY LIFE, AND TIES WITH EXTENDED FAMILY

Others (e.g., Axinn & Thornton 1992; Bennett, Blanc & Bloom 1988; Willis & Michael 1994) suggest that the relative instability of cohabitational unions makes them well suited to those who wish conjugality without long-term commitment. Childbearing implies long-term commitment, as does maintenance of ties with extended family. So we hypothesize that the propensity to marry rather than to cohabit is increased by values and attitudes that are favorable to marriage, childbearing, and maintenance of ties with relatives.

SEX-ROLE LIBERALISM

Traditional sex roles are differentiated, with women's roles centered on household and children, and men's roles centered on economic activity outside the home (England & Farkas 1986). Sex-role traditionalism refers to a preference for these differentiated roles, while sex-role liberalism indicates a preference for less differentiation between men's and women's roles, and less segregation of roles by sex into home- and work-centered activities. We hypothesize a negative effect of sex-role liberalism on the probability of marriage as opposed to cohabitation.

IMPORTANCE OF LEISURE

Marriage increases hours of employment for men and hours of housework for women, and childbearing makes heavy demands on the time of parents (Juster & Stafford 1985). So it appears that marriage provides less time for leisure pursuits than cohabitation. We hypothesize that attitudes and values that favor leisure time reduce the probability of marriage as opposed to cohabitation.

IMPORTANCE OF MONEY

Marriage is associated with childbearing and child rearing, which are exceptionally expensive. Marriage is also more associated than cohabitation with home ownership, and that too is very expensive. Even in the absence of children and home ownership, men may view marriage as a financial drain because marriage (but not usually cohabitation) gives spouses claims on each others' earnings, and women not only tend to earn less than men, but women's earnings tend to decline after marriage (England & Farkas 1986). In contrast, the same gender inequality in earnings may make marriage more attractive, or just less unattractive, to women who seek money, since the labor market provides females with opportunities for smaller average earnings than it gives to men and marriage tends to increase their income, while it decreases the income of males. So we

hypothesize that attitudes and values that favor an abundance of money tend to lower the probability of marriage as opposed to cohabitation. We hypothesize that this effect is larger for men than for women.

IMPORTANCE OF CAREER SUCCESS AND WORK

Because marriage increases men's average hours of work (Juster & Stafford 1985) and wages (Daniel 1992), marriage appears to be compatible with men's efforts to build successful careers. This tendency to increase hours of work after marriage has been found even for men who cohabited before marriage, suggesting that the marriage is more compatible than cohabitation with hours of work. Further, men who marry tend to have some of the burden of housework lifted by their wives, thereby enabling husbands to trade housework for time on the job. We therefore hypothesize that for men, attitudes and values that favor job and career success tend to raise the probability of marriage as compared to cohabitation.

For women, we expect the opposite effect of attitudes and values that favor job and career success. Marriage appears to be a hindrance to paid employment by wives and married women tend to work fewer hours for pay than their husbands, due at least in part to the disproportionate share of household and child-rearing tasks born by most wives (Coverman 1983). So we hypothesize that for women, attitudes and values that favor job and career success tend to lower the probability of marriage as compared to cohabitation.

CONTROLS FOR OTHER INFLUENCES ON UNION FORMATION

All our hypotheses are subject to the stipulation that relevant "other things" that affect union formation are held constant. These things are correlates of work, leisure, money, family, and sex-role attitudes and other values that affect nuptiality and household formation patterns, including race (Bumpass & Sweet 1989), sex, religion of upbringing (Thornton, Axinn & hill 1992), parental characteristics (Bumpass & Sweet 1989), years of schooling and school enrollment (Thornton, Axinn & Teachman 1994), employment, income, and age. Because most contemporary American religions encourage marriage and strongly discourage cohabitation, and because time spent in religious activities has been shown to affect the probability of cohabitation (Thornton, Axinn & Hill 1992), it is also appropriate to control for religious activity.

Data

We use data on 12,841 respondents who participated in the 1986 wave of the National Longitudinal Study of the High School Class of 1972 (NLS-72), which was first administered to a nationally representative sample of high school seniors in 1972, when respondents were about 18 years old. Follow-up survey waves were fielded in 1973, 1974, 1976, 1979, and 1986. The 1986 wave included extensive individual histories of marriages and cohabitations up to the age of approximately 32 years. Data from this fourteen-year span permit analysis of first union formation from late adolescence through the early thirties, a

"demographically dense" (Rindfuss 1991) time of life in which the vast majority of first coresidential unions are formed. To permit the discrete-time event history analyses with time-varying covariates that we describe below, the longitudinal NLS-72 data were reorganized into a file of person-year records, with one record for each person in each year, rather than a single record for each person in the sample. Because October is the reference month for many NLS-72 variables, each year begins in October and ends the following September.

MARRIAGE AND COHABITATION MEASURES

The 1986 wave of the NLS-72 gathered extensive retrospective marriage and cohabitational histories. These histories include the starting and ending dates of up to three marriages or "marriage-like relationships" lasting one month or more (85% of respondents reported at least one such union; 55 individuals reported more than three).[2] For each relationship, respondents were asked the date of the start of the relationship and whether they were married to the person when they started living together. For each respondent in each year between 1972 and 1986, we construct two dummy variables: the first is an indicator for the start of a first union during the year; the second is a dummy variable set to one if the union was a cohabitation and zero else.

ATTITUDE MEASURES

In each wave of NLS-72, respondents were given a list of activities and goals, and asked "How important is each of the following to you in your life? Very Important, Somewhat Important, or Not Important?" Six items on this list correspond to attitudes and values listed among our hypotheses: (1) "Finding the right person to marry and having a happy family life," (2) "Living close to parents and relatives," (3) "Being successful in my line of work," (4) "Being able to find steady work," (5) "Having lots of money," and (6) "Having leisure to enjoy my own interests."

Sex-role liberalism is measured on a scale obtained by factor analysis of responses to ten Likert scale questions on sex-role attitudes that appeared in the 1976 and 1979 survey waves. These items and their factor analysis are described in Appendix A. In both waves, the factor *liberalism of sex roles* correlates about -0.8 with the extent of agreement with the statement "It is better for everyone if the woman stays home and the man goes out to earn a living." A higher factor score indicates more acceptance of nonfamilial roles for women.

CONTROL VARIABLES

Control variables are listed in Table 1. Control variables are "fixed" if their values cannot vary between 1972 and 1986, and "time-varying" if their values can change during this time. Fixed controls include sex, religion in which the respondent was raised, race, and parental characteristics. Sex effects are controlled by performing all analyses separately for males and females. Religion is represented with dummy variables for Catholic religion and Other Christian religion.[3] A third dummy for any other religion, a category dominated by

Protestants and including no religion, is omitted from analyses to avoid multicollinearity. Race is represented with a dummy variable for black. Parental characteristics include the natural logarithm of the parents' annual income in 1972, the mean years of school completed by the respondent's parents, and a dummy variable equal to one if the respondent was raised in an intact family (living in a household with both natural parents in the senior year of high school).[4]

Time-varying controls include years of school completed, the natural logarithm of current earnings, a dummy variable equal to one if the respondent reported having had a child, and dummy variables for current school enrollment, current employment, and current membership in a church.[5] Age is held constant by performing separate analyses for 18-22-year-olds and for 22-31-year-olds, and by including variables for the calendar year and the calendar year squared.

Statistical Methods

Two key features of our hypotheses shape our choice of statistical method. First, the hypotheses concern the occurrence of the events marriage and cohabitation and that suggests the use of event history techniques. Second, our hypotheses are about attitudes. Attitudes can vary over time, and that variation should be recognized in the analysis to the extent possible. Although surveys can measure attitudes at specific times, and pairs of longitudinal survey waves can establish intervals during which attitudes change, surveys cannot identify the precise times between interviews when change occurs.[6] So we use a logistic regression discrete-time event history model. This technique approximates results from continuous-time proportional hazards models (Allison 1984; Yamaguchi 1991) while easily permitting covariates to vary over time and accommodating the unavailability of continuous time measures of attitudes and other key variables.

Our dependent variable involves three competing states — single, cohabiting, married — which makes analysis slightly more complicated than most two-state event history models. Following Allison (1984), Yamaguchi (1991), and Agresti (1990), we use the method of continuation-ratio logit models to examine these events. That is, we perform two logistic regression analyses: The first analysis is based on all person-years that begin with the respondent in a never-married, never-cohabited state. The dependent variable is the dummy variable for formation of a first coresidential union in the year. The second analysis is based on all person-years in which individuals formed a first marriage or first cohabitation. In this second analysis, the dependent variable is a dummy variable for cohabitation. This second model allows us to differentiate those who choose marriage from those who choose cohabitation.

For each person-year, time-varying variables are measured in the most recent previous year for which data are available. Using attitude measures from previous time intervals avoids the possibility that observed attitudes could be the result of residential union formation behavior in the interval under consideration. This lag reduces the reliability of attitude measures, we suspect, but that reduction would bias coefficients downward, making our hypothesis

TABLE 1: Definitions of Independent Variables

Name	*Definition*	*Range*	*Years Measured*
Marry	Importance of "finding the right person to marry and having a happy family life"	1-3	72-74, 76, 79
Close to parents	Importance of "living close to parents and relatives"	1-3	72-74, 76, 79
Success	Importance of "being successful in my line of work"	1-3	72-74, 76, 79
Steady work	Importance of "being able to find steady work"	1-3	72-74, 76, 79
Money	Importance of "having lots of money"	1-3	72-74, 76, 79
Leisure	Importance of "having leisure to enjoy my own interests"	1-3	73, 74, 76, 79
Liberalism of sex roles	Factor scores measuring sex-roles liberalism	-3.1-2.2	76, 79
Black	Indicator for black	0-1	
Catholic	Indicator for Catholic	0-1	72
Other Christian	Indicator for "other Christian"	0-1	72
Religion	Indicator for church membership	0-1	76, 79
Parental income	Natural log of parental income	8.4-10.3	72
Imputing income	Indicator for imputing a value on *parental income*	0-1	72
Parents' education	Mean of parents' educational attainments	10-18	72
Intact family	Respondent living with both natural parents in 1972	0-1	72
Educational attainment	Number of years of schooling completed	11-20	72-85
Student	Indicator for enrollment in school	0-1	72-85
Working	Indicator for employment	0-1	72-85
Child	Indicator for presence of a child or children	0-1	72-85
Income	Natural log of respondent's total income	0-11.9	72-85
Year	Year in pooled-interval of person-year record	72-85	72-85
Year square	Square of *year*		

tests conservative.[7] For two reasons, we perform separate analyses for person-years corresponding to ages 18 through 22 and 23 through 31. First, others find that union formation varies substantially by age (e.g., Goldscheider & Waite 1986; Waite & Spitze 1981). And, second, some attitude measures are available only at later ages: attitudes toward leisure was not included until the 1973 survey wave, and sex-role attitudes were not included until the 1976 wave. We present estimates of models with and without these variables for the later ages to allow comparison with models for the younger ages.

Logit coefficients lack an intuitively appealing metric, so we transform the logit coefficients for each attitude or value measure into the change in probability of union formation associated with a specified increase in that variable. For example, a two-unit change in the variables *marry, closepar, success, steady work, money,* or *leisure* is a change from reporting that the corresponding aspect of life is "not important" to reporting that it is "very important." These effects are nonlinear and nonadditive; we evaluate them at the average probability level for each subsample in which data are analyzed (Peterson 1985; Stolzenberg 1979).

Note that the two dependent variables in our analysis differ in an important way. The model of union formation estimates the effect of various attitudes and values on the probability of forming a first union during a year, given that one had not done so at the beginning of the year. Although most individuals have formed a first union by the age of 31, the proportion forming a union in any particular year is small, varying between 9 and 14%, according to age and sex (see Table 2). Thus, substantively modest effects on the probability of union formation in any year could cumulate over a number of years to a quite sizable impact on the outcome. These effects can be viewed as affecting the speed with which young men and women enter a first coresidential union.

The second set of models estimates the effect of attitudes and values on the type of first union chosen, whenever this choice is made. Clearly, one can decide whether to enter a union this year over a number of separate years, but one decides on the type of union — given that one is entered — only once. The mean proportion cohabiting in first union, given in Table 2, varies from 22 to 37%, according to age and sex.

Logit effects, of the type we describe above, are small when the mean probability of the relevant outcome is small, as is the case for the annual probability of union formation. The same coefficient would imply a much larger effect on an event, such as cohabitation versus marriage in first union, which has a mean probability between .25 and .75. Untransformed logit coefficients for key variables are presented in Appendix B and these allow comparison of the size of effects in subgroups having different mean probabilities of the outcome in question.

Results

Complete results of all logistic regression analyses appear in Appendix B. Table 2 gives the effects of variables directly related to our hypotheses, transformed into probability differences associated with a two-point or one standard deviation change in the independent variable, as appropriate.

TABLE 2: Effects of Attitudes toward the Family, Work, Money, Leisure, and Sex Roles on the Annual Probability of Forming a First Union, and Probability that the First Union Is a Cohabitation[a]

Independent variable	Males 18-22	Males 23-31	Males 23-31	Females 18-22	Females 23-31	Females 23-31
Effects on annual probability of forming first union						
Marry	0.106***	0.127***	0.131***	0.166***	0.163***	0.157***
Close to parents	-0.021***	-0.023**	-0.023**	-0.024***	-0.022*	-0.023**
Success	0.012	0.003	0.002	-0.019*	-0.014	-0.012
Steady work	0.015	0.016	0.005	-0.032***	-0.016	-0.018
Money	-0.023***	-0.026**	-0.029***	-0.008	-0.012	-0.015
Leisure	...	...	0.028*	...	...	0.026*
Liberalism of sex roles	...	...	-0.003	...	...	-0.008*
Mean annual probability	0.0890	0.1348	0.1362	0.1341	0.1303	0.1320
Person-years	19,356	15,403	13,772	18,552	13,017	11,965
Effects on probability that union is a cohabitation						
Marry	-0.148***	-0.177***	-0.176***	-0.184***	-0.265***	-0.226***
Close to parents	-0.044	-0.158***	-0.120***	-0.112***	-0.139***	-0.129***
Success	-0.068	-0.035	-0.057	0.009	0.203**	0.134*
Steady work	-0.034	-0.142***	-0.152***	-0.020	-0.057	-0.077
Money	0.018	0.094*	0.090*	0.057*	0.151**	0.137**
Leisure	...	...	0.198***	...	...	0.032
Liberalism of sex roles	...	...	0.093***	...	...	0.117***
Mean probability	0.2171	0.3370	0.3337	0.2235	0.3691	0.3699
Person years	1,723	2,077	1,876	2,488	1,696	1,579

[a] Based on logistic regression models reported in Appendix B. The effects of *marry, close to parents, success, steady work, money,* and *leisure* result from a two-point increase on the three-point scale in which 3 = "very important," 2 = "somewhat important" and 1 = "not important." The effect of *liberalism of sex roles* results from an increase of one standard deviation in the sex-role attitude factor score.

One-tailed: * $.01 < p < .05$ ** $.001 < p < .01$ *** $p < .001$

MARRIAGE AND FAMILY

Table 2 presents effects of a two-unit increase in the importance of "finding the right person to marry and having a happy family life" (*marry*); and "living close to parents and relatives" (*close to parents*). A two-unit change on these three-point scales is a change from a response of "not important" to "very important." The top panel, which presents the effects on the annual probability of forming a first union, shows that for both men and young women, favorable attitudes toward marriage increase the probability of union formation in a given year quite substantially. At the mean, an increase from "not important" to "very important" in *marry* raises the probability that a union is formed in any given year from 9% to 20% for men 18 to 22, and from 13% to 30% for women the same age.[8] Similar effects appear at later ages for men and women in both models examined. But, as hypothesized, if a union is formed, that same two-point increase in the importance of *marry* lowers the probability that the union is a cohabitation from 22% to 7% at younger ages for men and from 22% to just 4% at younger ages for women. The effects on union type are even larger at later ages — for women, there is a significant increase (at the 0.05 level) in the size of the effect of *marry* across the ages examined. Thus, the substantive effects of changes in the reported importance of marriage are quite large for both speed of entry into a first union and for the type of union chosen.

Note, however, that wanting to live close to parents and relatives slows the transition to first union for both men and women at all ages; a two-point increase in *close to parents* would decrease the probability of a first union in any given year by about 2%. For men 18 to 22, for example, this would mean a decline from 9% to 7% in the probability of entering a first union. In retrospect, it appears that this measure may reflect attachment to family of origin rather than an interest in family formation.

Table 2 also supports the hypothesis that among those who form first unions, persons who place high value on living close to parents and relatives (*close to parents*) are less likely to cohabit than to marry. However, these negative effects are not statistically significant for 18-22-year-old males and are one-half to one-third the size of corresponding effects of *marry* for women and 23-31-year-old men. These effects are sometimes substantively important, however; a two-point increase in the importance of living close to relatives would reduce the chances of cohabiting at first union by 11 to 14% for women at all ages and for men at later ages. Thus, for women from 18 to 22 years old an increase in importance of living close to relatives would reduce the probability of cohabitation rather than marriage from 22 to 11%, with similar changes for older women and older men.

MONEY AND WORK

The third through fifth rows of Table 2 present effects of *success* (importance of "being successful in my line of work"), *steady work* (importance of "being able to find steady work"), and *money* (importance of "having lots of money"). We hypothesized that men for whom career success is important would find the traditional male breadwinner role more comfortable than would men who do

not value career success for themselves. For men, *success* and *steady work* have very small, statistically insignificant effects on the probability of union formation. For men who do form a first union, both *success* and *steady work* decrease the chances that it is a cohabitation rather than marriage at both young and later ages, but only the effect of *steady work* for 23-31-year-olds is statistically significant. For these men a two-point increase in the importance of steady work would decrease the probability of cohabitation from 34% to 20%, a finding that is consistent with our hypothesis.

We also hypothesized that the importance of money would reduce the probability of marriage. Attitude toward money, for men, has effects that are statistically significant, negative, and modest in size; the effects in Table 2 imply that a two-point increase in the importance of money would decrease the annual probability of a first union for men from 9% to 7% at younger ages and from 13% to 11% at later ages. Table 2 shows that a two-point increase in the importance of money would increase the probability that a first union is a cohabitation from 34% to 43% for older men, as we hypothesized. However, *money* shows no effect on the probability of cohabitation at younger ages for men. Thus, the importance of money seems to reduce union formation for men and — at least among older men — to reduce the probability of marriage over cohabitation for those forming a union.

We hypothesize that women for whom personal career success is important would find the traditional wife role implied by marriage to be less attractive than would women for whom career success was not important. Career-oriented women would, we hypothesized, be less likely to enter a union than other women and more likely to cohabit if they did form a union. Table 2 shows negative effects of *success, steady work,* and *money* on the probability of first union formation for younger and older women, although these are all small and only two of the three variables reach statistical significance and then only at the younger ages. For women 18 to 22 years old, the importance of success at work and steady work reduce the probability of union formation. For these young women, an increase of two points in the importance of steady work decreases the annual probability that they form a union from 13% to 10%. A two-point increase in *success* decreases the annual probability from 13% to 11%. These effects are significantly different (at the 0.05 level) than the effects for men, which run in the opposite direction. We see no significant effect of *money* on women's union formation probabilities.

If women do form a union, we see effects of the type we hypothesized; the importance of career success increases the probability of choosing cohabitation. For 18-22-year-old women who do form a first union, the probability that the union is a cohabitation is unaffected by *success* and *steady work,* but *money* shows a small positive effect on the probability that the union is a cohabitation. A two-point increase in the importance of money implies an increase from 22% to 28% in the probability that the first union is a cohabitation. For 23-31-year-old women, a two-point increase in the importance of success at work raises the chances of cohabitation at first union from 37% to 57%, an effect that opposes the direction of and is significantly different from the effect for men at the same age. The same increase in the importance of money raises the probability of

cohabitation from 37% to 52%. These effects are large enough to be important substantively.

These results are generally consistent with our reasoning that men and women for whom traditional expectations about marriage fit well with their personal values are more likely to marry than others. Although the expected effects do not appear for all measures of values toward career or money, those results that do appear consistently show the expected sign and some are substantively important.

LEISURE

We hypothesized that because rather large differences appear in the leisure time of married versus unmarried men, those men to whom leisure time for their own pursuits is important may find marriage relatively unappealing for that reason. The sixth row of Table 2 presents effects of *leisure* for men and women 23 to 31 years old (this measure is not available at earlier ages). *Leisure* has small positive effects on the annual probability of union formation for both men and women. If a union forms, for women *leisure* has no significant effect on the probability that the union is a cohabitation. But for men who do form a union, a two-point increase in the importance of *leisure* increases the probability that the union is a cohabitation from 33% to 53%, which is quite consistent with our hypothesis and suggests important differences between marriage and cohabitation in expectations about allocation of leisure time.

SEX-ROLE ATTITUDES

We hypothesized that those men and women with relatively liberal views of sex roles would be attracted to cohabitation because of its typically less sex-linked division of labor than that associated with marriage. Recall that the measure of sex-role attitudes is available only at later ages. Table 2 shows small effects of sex-role liberalism on the annual probability of forming a first union for women but not for men. However, for both men and women who form a union we see fairly sizable effects of sex-role liberalism on the probability that the union is a cohabitation. A one-standard-deviation increase in sex-role liberalism would increase the probability of cohabitation versus marriage from 33 to 43% for men and from 37 to 49% for women. A comparison of the two models for women at later ages shows that some of the impact of *success* and *money* is mediated through sex-role attitudes. Thus, as might be expected, women to whom career success is important also hold relatively liberal views of sex roles. The results show, as we hypothesized, that men and women who reported relatively liberal views of sex roles prior to first union seem to take those views into account in deciding on the type of union to enter later.

Conclusions

This article argues that marriage and cohabitation provide those who enter them with different sets of expectations for a wide range of activities, circumstances, and social relations. We focus here on relations with extended family, leisure pursuits, employment, division of economic resources, and the division of household labor. We test and find support for the hypothesis that the choice

between cohabitation and marriage is affected by attitudes and values toward work, family, leisure time-use, money, and sex roles, as well as values and attitudes toward marriage itself.

For several reasons we think that our measures of attitudes and values are biased against showing effects on union formation and choice of marriage versus cohabitation. First, our measures of attitudes and values precede our observations of union formation by as much as seven years, especially at later ages. Attitudes and values may change over time and old attitudes tend to have a weaker relationship to current behavior than do current attitudes (McGuire 1985). Second, our measures of attitudes and values — evaluative research attitude measures — are the type least likely to show empirical relationships with behavior; they tell us what people think is good rather than what they are prone to do (Ajzen & Fishbein 1980; Stolzenberg 1994). Given this bias against finding effects and certainly against finding strong effects, we think that the consistent pattern of results, combined with the sizable substantive impact of some measures, provides support for our hypotheses.

At a glance, the similarity between cohabitation and marriage is apparent: they are both romantic coresidential unions. However, researchers faced with a large and persistent empirical association between premarital cohabitation and subsequent divorce have postulated that the cohabiting population is a "select" group — particularly in terms of having liberal values (Axinn & Thornton 1992; Lesthaeghe & Surkyn 1988; Thomson & Colella 1992). These liberal values, we suggest, represent more than a general political orientation. Rather, we argue that these values are indicators of preferences for a kind of union with distinctly different characteristics than marriage. Our results are consistent with the idea that cohabitation is attractive as an alternative to marriage not only because it is a tentative, nonlegal form of a coresidential union but, more broadly, because it accommodates a very different style of life.

A sizable proportion of cohabiting couples eventually marry. For these couples cohabitation may allow them to delay rather than avoid the assumption of marital roles. Certainly, some young adults avoid marital roles that they would currently find constraining or burdensome by avoiding any union. We argue here that cohabitation as an institution allows for flexibility and freedom from traditional gender-specific marital roles, at least temporarily.

Both men and women who reject the constraints and demands of the traditional marriage are more likely than others to choose informal unions. First, our results indicate that those who find marriage in and of itself less important as a life goal are more likely to enter into a cohabitational relationship in their first coresidential union. But even controlling for this, a broad array of attitudes — all associated with marital roles and expectations — affect union formation rates in our models. Beyond one's own attachment to marriage, for example, our results demonstrate that less attachment to one's parents and other kin also increases the propensity to cohabit. Not only can the family, much like a church (Thornton, Axinn & Hill 1992), reinforce pro-marriage values and discourage their ranks from cohabiting, but the "work of kinship" and family cohesion is, in itself, a part of marital roles (Di Leonardo 1987). Those who wish to distance themselves from their families may prefer to avoid the expectations of kin bonding that marriage often places upon them.

Work and the division of labor within the family is, of course, one of the most prominent normative elements of marriage. Our results indicate, then, that men who don't think that success at work or being able to find steady work for themselves is important are drawn disproportionately toward cohabitation, with its egalitarian pattern of responsibility for financial support. And women who value money and career success for themselves are also drawn to cohabitation, which may give them the freedom to pursue their goals unfettered by a need to make accommodations to husbands' careers. But very young women — those 22 years old and younger — tend to avoid any union if they value their own career success, perhaps to allow themselves the freedom to concentrate their energies on schooling and the transition to work, perhaps to retain the financial support of their parents.

We note that a focus on "having lots of money" consistently inhibits union formation, although this result occurs in different ways for men and women. A high value placed on money slows men's entry into any union, but for women, it inhibits marriage, not union formation. To the extent that the egalitarian ideal of cohabitation is tempered by the traditional expectations of financial responsibility in dating and marriage, this can be expected; a cohabiting man with money may feel compelled to pay for some expenses for his partner. On the other hand, this traditional assignment of financial responsibility had adverse consequences for women as well — historically, for example, when women married they often lost control of their own financial resources and property to their husbands. But women may feel less compelled to share financial assets with a cohabiting partner than with a spouse, inhibiting marriage by women who value money. Our results suggest that women who find having lots of money important do not tend to find that either living with or marrying a well-to-do man offers a route to this goal; perhaps finding such a man is more difficult than preparing for a career oneself.

The difference in the demands of cohabitation and marriage appear most strikingly in attitudes toward leisure time; other things equal, young men who highly value leisure time to pursue their own interests are much more likely than those less concerned with this issue to choose cohabitation later. Previous research suggests that marriage affects men's leisure quantitatively as well as qualitatively — not only do married men have less leisure time than single men (Shelton 1992), but they are much less likely to spend their leisure time away from their partners (Blumstein & Schwartz 1983). Our results indicate that those who want to preserve their own leisure time do not shun unions in general — rather they choose an informal union that may not carry the stringent requirements for sharing activities that accompany marriage.

Our results suggest that young men and women both see marriage as constraining. Marriage requires men to be successful at their career, hold a steady job, and give up their individual leisure pursuits. For young women, the burdensome demands avoided by some appear to revolve around giving up careers — or at least making them secondary to those of their husbands. For some young adults, the expectations associated with marriage fit well with their personal values. But other young adults appear to reject these goals — and the traditional sex roles that support them — but to not abandon the prospect of a coresidential romantic union altogether.

APPENDIX A: Item Loadings on Sex-Role Attitude Factor, 1976 and 1979[a]

Sex-Role Attitude Item	Loading 1976	Loading 1979
a. A working mother of preschool children can be just as good a mother as the woman who doesn't work.	-.3124	-.4011
b. It is usually better for everyone if the man is the achiever outside the home and the woman takes care of the home and family.	.7119	.7304
c. Young men should be encouraged to take jobs that are usually filled by women (nursing, secretarial work, etc.).	-.0084	-.0170
d. Most women are just not interested in having big and important jobs.	.4449	.4403
e. Many qualified women can't get good jobs: men with the same skills have less trouble.	-.0037	.0128
f. Most women are happiest when they are making a home and caring for children.	.6706	.6448
g. High school counselors should urge young women to train for jobs which are not held mainly by men.	-.0148	-.0498
h. It is more important for a wife to help her husband than to have a career for herself.	.6914	.7307
i. Schools teach women to want less important jobs.	.0198	.0673
j. Men should be given first chance at most jobs because they have the primary responsibility for providing for a family.	.5959	.6348

[a] After exploratory analyses that indicated that a factor structure with two dimensions fits the data well, we performed maximum likelihood factor analysis with a nonorthogonal promax rotation using SAS procedure factor. The two distinct dimensions produced by the factor analysis indicated that the first strongly related to views of traditional male-female roles in the family, and the second related to views of the appropriate means for correcting gender inequality. We estimated all models reported in this article with both factors, but only the first showed significant effects. In 1976, this single, weighted, rotated factor explains 42.8% of the variance in the items and has a Tucker-Lewis reliability coefficient of 0.92. In the 1979, this weighted, rotated factor explain 47.7% of the variance in the items and has a reliability coefficient of 0.93.

APPENDIX B: Logit Coefficients, Analyses of Annual Probability of Transition from Singlehood to a First Union, by Sex and Age Group

	Males			Females		
Independent variable	18-22 β (S.E.)	23-31 β (S.E.)	23-31 β (S.E.)	18-22 β (S.E.)	23-31 β (S.E.)	23-31 β (S.E.)
Marry	0.4530*** (0.643)	0.4104*** (0.0522)	0.4200*** (0.0567)	0.5087*** (0.0584)	0.5078*** (0.0572)	0.4920*** (0.0613)
Close to parents	-0.1430*** (0.0431)	-0.1080** (0.0397)	-0.1080** (0.0424)	-0.1118*** (0.0356)	-0.1040* (0.0429)	-0.1082** (0.0452)
Success	0.0687 (0.0710)	0.0129 (0.0602)	0.00656 (0.0637)	-0.0861* (0.0453)	-0.0670 (0.0650)	-0.0532 (0.0683)
Steady work	0.0859 (0.0600)	0.0650 (0.0524)	0.0199 (0.0552)	-0.1519*** (0.0448)	-0.0730 (0.0550)	-0.0841 (0.0575)
Money	-0.1596*** (0.0443)	-0.1220** (0.0408)	-0.1351*** (0.0437)	-0.0362 (0.0391)	-0.0575 (0.0455)	-0.0684 (0.0477)
Leisure	...	...	0.1100* (0.0513)	...	...	0.1043* (0.0576)
Liberalism of sex roles	...	...	-0.0225 (0.0324)	...	...	-0.0730* (0.0354)
Black	-0.3034** (0.1121)	-0.1348 (0.1059)	-0.1158 (0.1135)	-0.6059*** (0.0830)	-0.2955** (0.0922)	-0.2130* (0.0971)
Catholic	-0.2450*** (0.0615)	-0.0327 (0.0539)	-0.0242 (0.0568)	-0.1888*** (0.0516)	0.00116 (0.0618)	-0.0234 (0.0650)
Other Christian	0.1917** (0.0728)	0.1074 (0.0772)	0.0988 (0.0814)	0.0327 (0.0645)	-0.0797 (0.0844)	-0.0855 (0.0871)
Religiosity	...	...	0.1056* (0.0558)	...	...	-0.0963 (0.0589)

APPENDIX B: Logit Coefficients, Analyses of Annual Probability of Transition from Singlehood to a First Union, by Sex and Age Group (Continued)

	Males			Females		
Independent variable	18-22 β (S.E.)	23-31 β (S.E.)	23-31 β (S.E.)	18-22 β (S.E.)	23-31 β (S.E.)	23-31 β (S.E.)
Parental income	0.0340* (0.0184)	-0.00781 (0.0451)	0.00116 (0.0477)	0.0199 (0.0143)	0.0397 (0.0515)	0.0498 (0.0537)
Imputing income	0.0563 (0.0689)	-0.0491 (0.0637)	-0.0333 (0.0676)	-0.0951* (0.0525)	-0.0549 (0.0606)	-0.0561 (0.0637)
Parents' education	-0.0888*** (0.0146)	-0.00537 (0.0133)	-0.00104 (0.0140)	-0.1040*** (0.0123)	0.00642 (0.0152)	0.00768 (0.0158)
Intact family	0.00910 (0.0597)	0.1016 (0.0566)	0.0840 (0.0600)	-0.0674 (0.0503)	0.0534 (0.0637)	0.0655 (0.0661)
Educational attainment	0.0490 (0.0496)	0.0571*** (0.0132)	0.0562*** (0.0141)	0.1181** (0.0424)	0.0383* (0.0153)	0.0465** (0.0162)
Student	-0.6447*** (0.0607)	-0.0270 (0.0568)	0.00432 (0.0599)	-0.6578*** (0.0537)	0.1718* (0.0620)	-0.1625** (0.0643)
Working	0.3118*** (0.0698)	0.2363** (0.0805)	0.2589** (0.0851)	0.2349*** (0.0563)	0.2607** (0.0890)	0.2790** (0.0927)
Child	0.2758 (0.2488)	-0.3930** (0.1498)	-0.4377** (0.1616)	-0.0638 (0.1434)	-0.4961*** (0.1157)	-0.4827*** (0.1193)
Income	0.0914*** (0.0168)	0.0555*** (0.0143)	0.0435** (0.0147)	0.00945 (0.00933)	0.0398** (0.0136)	0.0443*** (0.0141)
Year	16.1892*** (3.4230)	2.0207** (0.7028)	1.6573* (0.7372)	4.6861* (2.4918)	1.2096 (0.7914)	1.0678 (0.8222)
$Year^2$	-0.1071*** (0.0229)	-0.0129** (.00436)	-0.0107** (0.00457)	-0.0309* (0.0167)	-0.00822 (0.00492)	-0.00737 (0.00511)
Intercept	-614.9*** (127.9)	-82.0123 (28.2682)	-68.7687* (29.6543)	-179.9* (92.8240)	-47.0819 (31.7984)	-42.3771 (33.0274)
-2 log likelihood	943.651***	214.153***	196.222***	743.845***	346.761***	339.779***
Degrees of freedom	19	19	22	19	19	22
Person years	19,356	15,403	13,772	18,552	13,017	11,965
Percent forming unions	8.90	13.48	13.62	13.41	13.03	13.20

One-tailed: * $.01 < p < .05$ ** $.001 < p < .01$ *** $p < .001$

APPENDIX C: Logit Coefficients, Analyses of Annual Probability That First Union Is a Cohabitation, by Sex and Age Group

Independent variable	Males 18-22 β (S.E.)	Males 23-31 β (S.E.)	Males 23-31 β (S.E.)	Females 18-22 β (S.E.)	Females 23-31 β (S.E.)	Females 23-31 β (S.E.)
Marry	-0.6568*** (0.1363)	-0.4911*** (0.1056)	-0.4909*** (0.1206)	-0.9704*** (0.1174)	-0.8101*** (0.1196)	-0.6261*** (0.1304)
Close to parents	-0.1397 (0.1020)	-0.4229*** (0.0839)	-0.3048*** (0.0928)	-0.4152*** (0.0845)	-0.3353*** (0.0868)	-0.3072*** (0.0937)
Success	-0.2302 (0.1607)	-0.0805 (0.1224)	-0.1353 (0.1368)	0.0244 (0.1210)	0.4129** (0.1378)	0.2751* (0.1471)
Steady work	-0.1051 (0.1377)	-0.3698*** (0.1035)	-0.4068*** (0.1148)	-0.0592 (0.0989)	-0.1279 (0.1121)	-0.1742 (0.1198)
Money	0.0502 (0.1023)	0.1988* (0.0841)	0.1912* (0.0946)	0.1527* (0.0907)	0.3073** (0.0945)	0.2800** (0.1024)
Leisure	...	...	0.4094*** (0.1138)	...	...	0.0669 (0.1215)
Liberalism of sex roles	...	...	0.3980*** (0.0719)	...	...	0.4799*** (0.0773)
Black	0.9123*** (0.2336)	0.4944* (0.2158)	0.6232** (0.2414)	0.3671* (0.1853)	0.0717 (0.1929)	0.1863 (0.2063)
Catholic	0.2113 (0.1421)	-0.0921 (0.1117)	-0.1643 (0.1230)	0.0510 (0.1178)	-0.0684 (0.1238)	-0.2179 (0.1332)
Other Christian	-0.2832 (0.1778)	0.1051 (0.1558)	0.2867* (0.1727)	-0.1302 (0.1513)	-0.0859 (0.1717)	0.0363 (0.1823)
Religiosity	...	...	-1.0350*** (0.1303)	...	...	-0.7484*** (0.1243)

APPENDIX C: Logit Coefficients, Analyses of Annual Probability That First Union Is a Cohabitation, by Sex and Age Group (Continued)

	Males			Females		
Independent variable	18-22 β (S.E.)	23-31 β (S.E.)	23-31 β (S.E.)	18-22 β (S.E.)	23-31 β (S.E.)	23-31 β (S.E.)
Parental income	-0.0165 (0.0430)	0.0785 (0.0921)	-0.00241 (0.1022)	-0.0414 (0.0332)	0.0676 (0.1059)	0.0178 (0.1137)
Imputing income	-0.2615 (0.1689)	-0.0279 (0.1322)	-0.0223 (0.1467)	-0.1529 (0.1250)	-0.1327 (0.1216)	-0.1580 (0.1316)
Parents' education	0.0972** (0.0339)	0.0769** (0.0279)	0.0759** (0.0306)	0.0901*** (0.0277)	0.0851** (0.0301)	0.0829** (0.0324)
Intact family	-0.3258** (0.1362)	-0.2628* (0.1140)	-0.1016 (0.1261)	-0.3443** (0.1127)	-0.3132* (0.1249)	-0.2466)* (0.1343)
Educational attainment	-0.0406 (0.1125)	-0.1193*** (0.0263)	-0.1509*** (0.0295)	-0.2349** (0.0995)	-0.0817** (0.0304)	-0.1120*** (0.0332)
Student	0.0226 (0.1422)	-0.0368 (0.1185)	0.0332 (0.1299)	-0.0107 (0.1227)	0.1511 (0.1253)	0.0826 (0.1345)
Working	-0.4073** (0.1538)	-0.3902*** (0.1647)	-0.4439** (0.1788)	-0.1699 (0.1303)	0.0195 (0.1830)	0.0622 (0.1988)
Child	-0.3244 (0.5492)	0.00476 (0.3075)	0.0715 (0.3437)	0.7994** (0.2861)	0.7691** (0.2424)	0.8242*** (0.2578)
Income	-0.0428 (0.0336)	-0.0111 (0.0262)	-0.0157 (0.0280)	0.0569** (0.0220)	0.00823 (0.0262)	-0.00296 (0.0282)
Year	-8.2043 (8.3165)	1.8026 (1.4547)	2.1726 (1.5902)	12.8088* (6.3370)	0.4464 (1.6085)	-0.3428 (1.7445)
Year^2	0.0554 (0.0556)	-0.0104 (0.00902)	-0.0125 (0.00986)	-0.0857* (0.0424)	-0.00241 (0.01000)	0.00257 (0.0108)
Intercept	305.5 (310.7)	-74.2864 (58.5468)	-90.1335 (64.009)	-474.6* (236.3)	-19.2510 (64.5812)	12.4622 (70.0175)
-2 log likelihood	105.838***	169.001***	303.582***	212.915***	141.388***	234.266***
Degrees of freedom	19	19	22	19	19	22
Person years	1,723	2,077	1,876	2,488	1,696	1,579
Percent forming cohabs	21.71	33.70	33.37	22.35	36.91	36.99

One-tailed: * .01 < p < .05 ** .001 < p < .01 *** p < .001

Notes

1. We think that the processes that underlie the transition from a cohabitation to marriage, or the transition from a disrupted marriage into cohabitation, may differ in important ways from the processes that determine the choice between cohabitation and marriage for the first union. We examine only the effects of attitudes on the first union.

2. If cohabiting respondents did not characterize their relationship as "marriage-like," they were not coded as cohabiting. We doubt that it was a substantial source of error in the NLS-72. To the extent that this error did occur, it would bias our analyses against our hypotheses that attitudes and values affect the probability of union formation and cohabitation.

3. Possible responses to the religion question were: Protestant; Roman Catholic; other Christian Religion; Jewish; other or Eastern religion; and no religion. The other Christian category appears to be dominated by fundamentalists.

4. Parental income was missing for about a fifth of the respondents. For almost all cases with missing data, we were able to calculate a predicted income based on parental education, father's occupational status, marital history, race and ethnicity, urbanicity and region of the country. We then added an error term to the predicted values so that the variance of the income variable was not changed by imputation (cf. Little & Rubin 1987). We also included a dummy variable to flag this imputation.

5. In 1972, when the respondents were still in high school, the questionnaire did not include a measure of the respondent's income. The base-year survey did, however, include a measure of number of hours a week worked at a job. For the present analysis, we used a rescaled measure of hours of work as a proxy for income in 1972. We suggest that while this technique may be problematic for adults, it is probably acceptable for high school students, since most of the entry-level jobs available to them would be paid by hourly rates at or near the minimum wage. To evaluate whether this imputation procedure made substantial differences in our models, we reestimated all models with and without income measurements. The coefficients for the other variables were virtually unchanged. We also reestimated our models excluding respondents who formed unions before October 1973 (that is, those for whom we did not have an actual income measurement). This did tend to decrease the coefficient for income. Thus, caution should be used when interpreting the coefficient for income in the youngest age group in our analysis, since our imputation procedure has clearly lowered the reliability of that variable.

6. This limitation is caused by the invalidity and unreliability of retrospective survey questions. Except under unusual circumstances, one cannot ascertain the time at which an attitude changed by asking respondents when those changes occurred. For example, imagine asking a person to name the day or month when he or she decided that living close to relatives changed from being "very important" to being merely "important." In contrast, one can use retrospective questions to ascertain the dates of significant life events, such as marriage or cohabitation.

7. Note that the time-lag for variables measured only in survey years, such as the attitudinal items, can be longer than a year. The time-lag is particularly long for the later years included in our analysis, as the last two waves were more than five years apart. Since more remote measurements of attitude are less reliable predictors of behavior (McGuire 1985), we expect that our results may understate the importance of attitudes on union formation, especially in the later years of young adulthood.

8. We calculate these changes in effects by adding the percentage change associated with a two-point increase in the independent variable of interest — in this case *marry* — to the mean annual probability that a union is formed for the relevant age and sex group. For all variables that appear in models for older and younger ages, we use percentage changes for the comparable model.

References

Agresti, Alan. 1990. *Categorical Data Analysis.* Wiley.

Ajzen, Icek, and Martin Fishbein. 1980. *Understanding Attitudes and Predicting Social Behavior.* Prentice-Hall.

Allison, Paul D. 1984. *Event History Analysis: Regression for Longitudinal Data.* Sage.

Axinn, William G., and Arland Thornton. 1992. "The Relationship between Cohabitation and Divorce: Selectivity or Causal Influence?" *Demography* 29:357-74.

Bachrach, Christine A. 1987. "Cohabitation and Reproductive Behavior in the U.S." *Demography* 24:623-37.

Becker, Gary. 1991. *A Treatise on the Family.* Harvard University Press.

Bennett, Neil G., Ann Klimas Blanc, and David E. Bloom. 1988. "Commitment and the Modern Union: Assessing the Link between Premarital Cohabitation and Subsequent Marital Stability." *American Sociological Review* 53:127-38.

Berk, Sarah Fenstermaker. 1985. *The Gender Factory: The Apportionment of Work in American Households.* Plenum Press.

Bielby, William T., and Denise D. Bielby. 1992. "I Will Follow Him: Family Ties, Gender-Role Beliefs, and Reluctance to Relocate for a Better Job." *American Journal of Sociology* 97:1241-67.

Blair, Sampson Lee, and Daniel T. Lichter. 1991. "Measuring the Division of Household Labor: Gender Segregation among American Couples." *Journal of Family Issues* 12:91-113.

Blumstein, Philip, and Pepper Schwartz. 1983. *American Couples: Money, Work, Sex.* William Morrow.

Booth, Alan, and David Johnson. 1988. "Premarital Cohabitation and Marital Success." *Journal of Family Issues* 9:255-72.

Brines, Julie, and Kara Joyner. 1993. "The Ties That Bind: Principles of Stability in the Modern Union." Paper presented at Population Association of America in Cincinnati, Ohio.

Bumpass, Larry L., and James A. Sweet. 1989. "National Estimates of Cohabitation." *Demography* 26:615-26.

Bumpass, Larry L., James A. Sweet, and Andrew Cherlin. 1991. "The Role of Cohabitation in Declining Rates of Marriage." *Journal of Marriage and the Family* 53:913-27.

Caplow, Theodore, Howard M. Bahr, and Bruce A. Chadwick. 1983. *All Faithful People: Change and Continuity in Middletown's Religion.* University of Minnesota Press.

Clarkberg, Marin. 1994. "The Socioeconomic Determinants of Cohabitation." Paper presented at the meetings of the Population Association of America, Miami, May.

Coverman, Shelly. 1983. "Gender, Domestic Labor Time, and Wage Inequality." *American Sociological Review* 48:623-37.

Daniel, Kermit. 1992. "Does Marriage Make Men More Productive?" Working Paper 92-2, Economics Research Center, NORC and University of Chicago.

DeMaris, Alfred, and William MacDonald. 1993. "Premarital Cohabitation and Marital Instability: A Test of the Unconventionality Hypothesis." *Journal of Marriage and the Family* 55:399-407.

Di Leonardo, Micaela. 1987. "The Female World of Cards and Holidays: Women, Families, and the Work of Kinship." *Signs* 12:440-53.

England, Paula, and George Farkas. 1986. *Households, Employment, and Gender: A Social, Economic and Demographic View.* Aldine.

Glick, P.C., and G.B. Spanier. 1980. "Married and Unmarried Cohabitation in the United States." *Journal of Marriage and the Family* 42:19-30.

Goldscheider, Frances K., and Linda J. Waite. 1986. "Sex Differences in Entry into Marriage." *American Journal of Sociology* 92:91-109.

Hochschild, Arlie. 1989. *The Second Shift.* Avon.

Juster, Thomas F., and Frank P. Stafford (eds.). 1985. *Time, Goods, and Well-Being.* Institute for Social Research.

Lesthaeghe, R., G. Moors, and L. Halman. 1993. "Living Arrangements and Values among Young Adults in the Netherlands, Belgium, France and Germany, 1990." Paper presented at the Population Association of America, Cincinnati, April.

Lesthaeghe, Ron, and John Surkyn. 1988. "Cultural Dynamics and Economic Theories of Fertility Change." *Population and Development Review* 14:1-45.

Little, Roderick J.A., and Donald B. Rubin. 1987. *Statistical Analysis with Missing Data.* Wiley.

McGuire, W.J. 1985. "Attitudes and Attitude Change." Pp. 223-346 in *Handbook of Social Psychology.* Vol. 2, 3d ed., edited by Gardner Lindzey and Eliot Aronson. Random House.

Parsons, Talcott, and Robert F. Bales. 1955. *Family, Socialization and Interaction Process.* Free Press.

Peterson, Trond. 1985. "A Comment on Presenting Results from Logit and Probit Models." *American Sociological Review* 50:130-31.

Rindfuss, Ronald R. 1991. "The Young Adult Years: Diversity, Structural Change, and Fertility." *Demography* 28:493-512.

Rindfuss, Ronald R., and Audrey VandenHeuvel. 1990. "Cohabitation: A Precursor to Marriage or an Alternative to Being Single?" *Population and Development Review* 16:703-26.

Shelton, Beth Ann. 1992. *Women, Men and Time: Gender Differences in Paid Work, Housework and Leisure.* Greenwood Press.

South, Scott J., and Glenna Spitze. 1994. "Housework in Marital and Nonmarital Households." *American Sociological Review* 59:327-47.

Stolzenberg, Ross M. 1979. "Measurement and Decomposition of Causal Effects in Nonlinear and Nonadditive Models." Pp. 459-88 in *Sociological Methodology.* Jossey-Bass.

____. 1994. "Educational Continuation by College Graduates." *American Journal of Sociology* 99:1042-77.

Stolzenberg, Ross M., Mary Blair-Loy, and Linda J. Waite. 1995. "Religious Participation in Early Adulthood: Age and Family Life Cycle Effects on Church Membership." *American Sociological Review* 60:84-103.

Teachman, Jay D., and Karen Polonko. 1990. "Cohabitation and Marital Stability in the United States." *Social Forces* 69:207-20.

Teachman, Jay D., Jeffrey Thomas, and Kathleen Paasch. 1991. "Legal Status and the Stability of Coresidential Unions." *Demography* 28:571-86.

Thomson, Elizabeth, and Ugo Colella. 1992. "Cohabitation and Marital Stability: Quality or Commitment." *Journal of Marriage and the Family* 54:259-67.

Thornton, Arland, William G. Axinn, and Daniel H. Hill. 1992. "Reciprocal Effects of Religiosity, Cohabitation, and Marriage." *American Journal of Sociology* 98:628-51.

Thornton, Arland, William G. Axinn, Jay D. Teachman. 1994. "The Influence of Educational Experiences on Cohabitation and Marriage in Early Adulthood." Paper presented at the meetings of the Population Association of America, Miami, May.

Waite, Linda J., and Glenna D. Spitze. 1981. "Young Women's Transition to Marriage." *Demography* 18:681-94.

Willis, Robert J., and Robert T. Michael. 1994. "Innovation in Family Formation: Evidence on Cohabitation in the U.S." Pp. 9-45 in *The Family, The Market and the State in Aging Societies,* edited by John Ermisch and Naohiro Ogawa. Clarendon Press.

Yamaguchi, Kazuo. 1991. *Event History Analysis.* Sage.

Yamaguchi, Kazuo, and Denise B. Kandel. 1985. "Dynamic Relationships between Pre-marital Cohabitation and Illicit Drug Use: An Event-History Analysis of Role Selection and Role Socialization." *American Sociological Review* 50:530-46.

Migrant Remittances, Labor Markets, and Household Strategies: A Comparative Analysis of Low-Income Household Strategies in the Caribbean Basin*

JOSE ITZIGSOHN, *Brown University*

Abstract

This article analyses the relationship between remittances and the subsistence strategies of low-income households in four capital cities of the Caribbean Basin. The switch to an export-oriented model of development implemented in response to the economic crisis of the beginning of the 1980s deteriorated the labor market conditions in the region, making it difficult to subsist on local incomes. The findings point to the importance of migrant remittances on urban low-income households in terms of the number of households receiving remittances, their importance for the household income, and their effects on labor market participation of the head of the household, as well as additional members. They suggest that remittances are a rational strategy followed by households confronting very difficult labor market conditions.

This article analyzes the effects of migrant remittances on the life and economic activities of urban low-income households in the Caribbean Basin. During the last two decades, labor markets have become internationalized. An increasing number of people have undertaken long journeys and border crossings to find their livelihood in foreign lands, while at the same time helping the relatives they left behind. The Caribbean is one of the regions where migration and

* *The research reported herein was conducted with the support of a grant from the Andrew Mellon Foundation. The project was a collaborative effort between institutions and research teams in each of the countries studied, under the direction of Professor Alejandro Portes of The Johns Hopkins University. The project directors and institutions involved were as follows: Mario Lungo, Central America University Confederation (CSUCA), Costa Rica; Wilfredo Lozano, Latin American Faculty of Social Science (FLACSO), the Dominican Republic; Juan Pablo Perez Sainz, FLACSO, Guatemala; Sabine Manigat, National University of the Republic, Haiti; and Derek Gordon and Patricia Anderson, University of the West Indies, Jamaica. David Lindstrom, George Wilson, Mahua Sarkar, Roberto Gutierrez, and two anonymous reviewers offered generous comments that greatly improved the article. My gratitude to them. The responsibility for the analysis and the content, however, is only mine.* *Direct correspondence to Jose Itzigsohn, Department of Sociology, 112 George St., Box 1916, Brown University, Providence, RI 02912.*

remittances have become extremely important to the subsistence of the low-income sectors. This work examines the effects of remittances on the subsistence strategies of urban households in four capital cities of the region: Santo Domingo, Guatemala City, Kingston, and Port-au-Prince.

Studies of development and remittances tend to focus either on the effects of remittances on the balance of payments or on their "welfare effect" on the population of the labor-sending country (Lundahl 1985; Martin 1991; Straubhaar 1992). Recently, a number of studies have examined the impact of remittances on the development of a small business sector linked to migration in the country of origin (Lopez & Seligson 1991; Portes & Guarnizo 1991). A number of researchers have focused on the effects of remittances on household income and labor market behavior (Funkhauser 1992; Massey et al. 1987; Selby & Murphy 1982), and some of these view migration and remittances as a rational household economic strategy. This article shares this approach in that it analyzes the demographic and socioeconomic characteristics of households that receive remittances and focuses on the importance of remittances for the household income and labor market insertion of household members. The comparative approach helps to distinguish those household processes that are common to the countries of the Caribbean Basin from those that are particular to each country.

An important contribution of this study is to situate the influence of remittances on household strategies within the framework of the socioeconomic crisis that affected Latin America and the Caribbean in the 1980s and the restructuring of the economies and urban labor markets that followed. Decline in local wage opportunities has provided a strong incentive for international migration. The money that migrants send back home is likely to become an important factor in the household budget (Funkhauser 1992; Kaimowitz 1990; Tardanico 1993).

Remittances and Development

A large number of studies on the effects of remittances on development focus on how remittances are spent and the effect of that spending on the economy of the sending country or region. There is no consensus on what those effects are, or how best to evaluate them. The debate centers around the empirical evidence showing that remittances are spent mostly on basic subsistence needs, and after those are fulfilled, on housing improvement and eventually land purchase (Appleyard 1989; Guarnizo 1992; Stahl & Arnold 1986).

Some scholars, mostly economists working from a neoclassical perspective, see remittances as having a positive effect on the development of the labor-sending area. On the macroeconomic level, the argument in favor of remittances asserts that they constitute a much-needed source of foreign currency. Remittances contribute to leveling the balance of payments and pay a large proportion of the import bill of the labor-sending country. Moreover, these scholars claim that despite being spent on consumption, remittances provide an important incentive for the local production of goods and services. Remittances

also improve both income distribution and the quality of life of large segments of the population (Arnold & Stahl 1986; Keely & Tran 1989).

Other scholars, mostly social scientists working from a historical-structural approach, see remittances as increasing the social inequalities in the labor-sending countries. Moreover, remittances are at best an unreliable source of foreign currency, since they depend heavily on the continuation of the labor migration process, thus increasing the dependency of the labor-sending countries on the willingness of the labor-receiving countries to keep receiving migrants (Birks & Sinclair 1980; Kritz & Keely 1981). Remittances do increase the purchasing power of certain segments of the population, changing their consumption patterns to include more imported goods, and creating a demonstration effect among nonmigrant households. Thus, according to this argument, remittances increase inequality, undermine the balance of payments (through increased demand for imported goods), and have inflationary effects. They are also said to cause an increase in land speculation and prices and a decrease in agricultural production (Bray 1987; Grasmuck & Pessar 1990; Martin 1991).

The empirical evidence regarding remittances is inconclusive. Researchers have found evidence of both beneficial and detrimental effects of remittances (Keely & Tran 1989; Grasmuck & Pessar 1991). Straubhaar (1992) analyzes the effects of labor migration and remittances on Turkey during the 1960s and 1970s and finds that at different times and in different regions of the country, remittances have had all the effects described above. A new wave of studies coming from the field of economic sociology are less deterministic in their predictions. These studies see the effect of remittances as dependent on the social capital and social networks of those who migrate and those who remain at home (Guarnizo 1992, Massey et al. 1987). Empirical studies from this perspective have focused so far on the effect of remittances on the development of small businesses in the home country. Under the changes brought on by structural adjustment policies, small businesses have become an important source of employment and an avenue for upward mobility. These studies find that in some cases, the existence of networks among migrants and their relatives at home may help in the development of a sector of small businesses linked to migration (Guarnizo 1992; Portes & Guarnizo, 1991).

The emphasis of the present study is placed on the relationship between remittances and the strategies that low-income households employ for subsistence and for upward mobility. Among the low-income population of the cities in the periphery, households, and not individuals, are the main economic unit. Women's wages are usually too low to enable female-headed households to live on one income. Men's earnings are higher, but they seldom allow the household to achieve desired patterns of consumption and recreation. It is the pooling of household resources that allows the low-income population to "make it" in cities ravaged by underemployment and unemployment (Roberts 1989). Households can try to cope with a deteriorating economic situation by sending additional household members into the labor market, decreasing their budget expenses, or allowing the migration of one or more of the members of the household. These decisions are made according to the particular structure of opportunities in each city, the timing of the economic cycle, and the life cycle of

the household (Grasmuck & Pessar 1991; Massey et al. 1987; Selby & Murphy 1982).

To consider households as a resource-pooling unit is not denying the power and interest differentials that exist along gender and age lines within the household. Household strategies are sometimes the result of consensual decisions, but often are imposed by some household members against the opinions of other individuals within the household. Households encompass ties of solidarity and affection, as well as points of tension and conflict (Beneria 1992; Safa & Antrobus 1992). At all events, households are important decision-making units. This work makes abstraction of intrahousehold conflicts and focuses on household economic strategies and their responses to remittances.

The main claim of this article is that remittances are a rational and important strategy for subsistence and mobility under deteriorating labor market conditions. In order to examine this claim, this study analyses three topics that are central to an understanding of the relationship between remittances and household dynamics in the new labor market conditions:

(1) What are the main characteristics of the households that receive remittances? Previous research has found that those households that could afford sending members abroad in order to receive remittances were the better off among the low-income sector households (Grasmuck & Pessar 1990; Selby & Murphy 1992). We claim that the economic crisis of the 1980s caused an expansion of the migration-remittances strategy to include the lower socioeconomic strata.

(2) What is the importance of remittances for household income? How important are remittances compared to other determinants of household income, such as the income of the head of the household or the presence a secondary income? Previous research has emphasized the importance of remittance income for low-income households (Funkhauser 1992; Massey et al. 1987; Selby & Murphy 1982). We claim that in the context of the new export-oriented development policies, remittances have become a fundamental source of income for lifting households out of poverty.

(3) What is the effect of remittances on the labor force participation of the head and other members of the household? Selby and Murphy (1982) found that in Mexico in the late 1970s, migrant-sender households sent more people into the labor force than nonsender households. Funkhauser (1992), on the other hand, found that during the 1980s, remittances allowed people to withdraw from the labor market in El Salvador, thereby relieving unemployment pressures. We claim that in the context of deteriorating labor markets, remittances allow households to send fewer people into the labor force, and thereby help reduce unemployment and underemployment.

The examination of these questions is important in order to understand why households in peripheral cities take upon themselves the cost and the strain of sending members abroad. This study also may contribute to understanding why migration persists in spite of an increasingly hostile reception in the core countries.

Data and Method

The data for this study were gathered in a survey conducted in four capital cities of the Caribbean Basin in the fall of 1991. This area was selected for the present study because during the previous decades it has been a major "labor export" area. The debt crisis and policies of structural adjustment enacted during the 1980s only served to increase the migration flows from the area (Portes & Grosfoguel 1994; Sassen 1988). The countries selected for the study were Jamaica, the Dominican Republic, Haiti, and Guatemala.[1] The four share a number of characteristics. First, they are all small countries with a high concentration of their population in the capital city. The four are beneficiaries of the Caribbean Basin Initiative, and their economies are highly dependent on the exports of a few commodities, mainly to the U.S. market.

The economic crisis of the 1980s brought about a shift in development policies. In order to balance their trade deficits and meet their debt obligations, the Latin American and Caribbean countries switched from an inward-oriented development strategy towards an export-oriented one, and implemented World Bank-inspired structural adjustment policies. These policies entailed an abrupt opening of formerly protected economies and their exposure to world market pressures, as well as the reduction of state intervention in the economy — i.e., a reduction of the fiscal deficit, trade and financial deregulation, as well as extensive privatization (Beneria 1992; Ghai & Hewitt de Alcantara 1990; Portes, Itzigsohn & Dore 1994).

The new export-oriented policies have had strong effects on urban labor markets. Formal employment in large firms decreased, public employment diminished, and public sector wages deteriorated. Open unemployment rose to high levels. Fewer jobs were created, and the new jobs were of a lesser quality than those created in the previous period (PREALC 1987). The countries of the region attempted to take advantage of the windows of opportunity opened by the Caribbean Basin Initiative, which prompted a rapid growth in assembly manufacturing in export processing zones (EPZ) or in "maquila" firms.[2] These industries, however, are based on the comparative advantage offered by the low cost of labor.

Table 1 shows both the depth of the crisis the area suffered during the 1980s and the importance that workers' remittances had on the economies of these countries. Guatemala and the Dominican Republic experienced a decline of GNP per capita throughout the decade. Jamaica underwent a recuperation and growth of GNP per capita in the second half of the 1980s, but had suffered a severe reduction during the first half. In Haiti, this indicator grew consistently, but it is still the lowest in the entire hemisphere, less than half that of the Dominican Republic. The table also shows the growing importance of workers' remittances as a source of foreign exchange. The size of remittances as a proportion of the export of goods and services has grown during the 1980s. In Jamaica and the Dominican Republic, this proportion declined toward the end of the decade, but is still higher than at the beginning of the decade.[3]

The four countries display differences in their levels of economic development and in their political system. They include the poorest nation of the hemisphere (Haiti), one of the most industrialized countries in Central America

TABLE 1: Evolution of GNP Per Capita and Remittances during the 1980s[a]

	GNP per Capita ($U.S.)	Workers' Remittances ($U.S. Millions)	Workers' Remittances as Percentage of Exports of Goods and Services
Jamaica			
1980	1,130	30,6	3.55
1985	920	92,3	7.28
1990	1,578	140,0	6.01
Haiti			
1980	260	106,4	34.47
1985	320	95,8	27.98
1990	370	117,0	47.10
Dominican Republic			
1980	1,090	183,1	13.9
1985	800	242,0	18.0
1990	820	314,8	15.6
Guatemala			
1980	1,130	—	—
1985	1,210	—	—
1990	910	106,6	6.7

[a] Source: World Bank, 1993

(Guatemala), and the country that has developed the largest EPZs in the region (the Dominican Republic). Politically they vary, including a stable democratic parliamentary system (Jamaica), a presidential democracy (the Dominican Republic), a fragile emerging democratic regime (Haiti), and an unstable democracy plagued by civil war and human rights violations (Guatemala). These political and economic differences are important to the comparative aspect of this study. Common findings registered in different settings indicate a more general trend than those observed in similar contexts. At the same time, if we find differences in the relationship between remittances and household strategies in the different cities, we can attribute these differences to the different characteristics of the countries (Przeworski & Teune 1970). The countries vary also in their migration history. A brief description of migration from each of the four countries is necessary to understand how differences between the migration processes from each country may affect household strategies in each city.

The destination of the Dominican migrants is mainly the U.S. This migration began after the invasion of the Dominican Republic by U.S. troops in 1965, initiated as a way to alleviate social and political pressures on the island. Until the 1980s, migrants were mainly from urban lower-middle-class origins. During the 1980s, due to the worsening of economic conditions on the island and the presence of a Dominican community in the U.S., migration is reported to encompass all social strata, from urban poor to professionals (Grasmuck & Pessar 1991; Guarnizo 1992).

Until the 1980s there were two types of migration from Haiti: first, there was the migration of laborers to the sugar cane fields of the Dominican Republic or the Bahamas; second, there was a migration of middle-class political refugees to Canada, France, and the U.S. During the 1980s, the economic crisis and the presence of a community of migrants in the U.S. prompted large numbers of impoverished Haitians to set sail on precarious boats in an attempt to reach Florida (Portes & Grosfoguel 1994).

Jamaican migration until the 1960s was directed mostly toward Great Britain. In the 1960s the British government erected barriers against migration from the Commonwealth countries, while labor recruiters began to bring workers from the West Indies to the sugar plantations of Florida. These two developments caused the tide of Jamaican migration to turn to the U.S. At the same time, Jamaican professionals also received incentives to migrate to the U.S. (Portes & Grosfoguel 1994).

Guatemala before the 1980s was not a labor-exporting country. During the 1980s, however, two important migration movements took place. First, civil war triggered a large migration from the rural areas to southern Mexico and Belize. Second, there was an economically motivated migration of urban origin headed for the U.S. (Hamilton & Stoltz Chinchilla 1991), a migration that is reflected in this study.

Data collection for this study was conducted by local research teams directed by senior social scientists. Budgetary limitations prevented the organization of city-wide representative surveys. Instead, the local teams designed a two-stage survey with a view to capturing the characteristics of the "popular" sectors in each city. In the first stage, the local research team selected a number of neighborhoods in each city[4] in an attempt to capture the variety of middle-low-income and low-income groups. The areas ranged from those inhabited mainly by public employees and formal manufacturing workers to shantytowns inhabited mainly by informal workers. In stage two, stage households were selected randomly within each neighborhood. In each household the head was interviewed.[5] Thus, the results of the survey cannot be generalized to the whole population of the city. Nevertheless, the findings provide important evidence concerning the economic behavior of middle-low-income and low-income sectors households.

In Santo Domingo and in Guatemala City, the research teams conducted 400 interviews. The survey encompassed four neighborhoods in the former city, and three in the latter. In Kingston, the sample comprised 800 cases due to the attempt to cover both the low-income population of the city itself and the adjacent suburbs. In Port-au-Prince, a survey of three neighborhoods was carried out. The plan was to sample 400 cases, but the military coup that ousted

President Aristide made completion of the survey impossible. The researchers, however, managed to complete 300 interviews, most of them before the coup.

The same questionnaire was applied in each city to insure comparability of results.[6] The questionnaire included items on demographic and socioeconomic issues such as marital and migratory status, number of children, education, occupation, income, members of the household in the economically active population, and ownership of the house. It also included questions about migrant relatives, their whereabouts, whether they send remittances or not, and the monthly amount sent. This information elucidates the economic strategies of the households in the cities included in this study.

Although the data used in this article are cross-sectional, they help us to understand the process of change that affected the Caribbean Basin during the 1980s. As shown in Table 1, the importance of remittances has greatly increased during the 1980s in all the countries included in this study. A comparative analysis of the effects of remittances on middle-low-income and low-income households in the context of deteriorating labor markets can provide an explanation for the trends shown in Table 1. Comparison with previous research will provide additional strength to our conclusions.

Characteristics of Households Receiving Remittances

How extensive are migration and remittances among middle-low-income and low-income households in the four cities? The survey asked whether the respondent has relatives abroad, the relationship of the relatives abroad to the head of the household, and the location of those relatives. The survey also asked whether the household receives help from relatives or friends abroad, and the monthly amount of this aid. The survey did not ask, however, who among those relatives sends remittances nor the country of origin of these remittances. Table 2 presents general findings about the importance of migration and remittances to our sample.

The first row of Table 2 shows that in each of the four cities, over half of the households have relatives abroad. Haiti and the Dominican Republic are the two countries where middle-low-income and low-income urban migration seem to be most pervasive. The second and third rows show that a large proportion of the relatives abroad are close relatives and that most of the migration households have relatives in the U.S. The last row shows that the households that receive remittances constitute a considerable portion of the households with relatives abroad. Guatemala exhibits the lowest proportion of households with remittances, but as mentioned above, urban migration from Guatemala is a relatively new phenomenon.

Table 3 presents the main demographic characteristics of households that receive remittances. The first two variables — gender and age — are individual characteristics of the head of household. These two variables indicate that women and older heads of household receive remittances more than other heads of household. However, these characteristics are significant only in Jamaica and Guatemala. In the Dominican Republic, only age is a significant variable, and in Haiti none of those characteristics is significant.

TABLE 2: Migration Characteristics of the Sample in the Four Cities

	Percentages			
	Jamaica	Haiti	Dominican Republic	Guatemala
Households with relatives abroad	59.3	69.0	76.7	54.7
Households with direct relatives abroad[a]	48.4	21.7[b]	53.6	36.8
Households with relatives in the U.S.	48.7	61.3	67.7	51.3
Households that receive remittances	19.4	40.3	26.8	14.5
Total	792	300	403	400

[a] Direct relatives are parents, sons and daughters, brothers and sisters, and spouse.
[b] In the case of Haiti, the low number of direct relatives may be due to the fact that brothers and sisters were not coded separately but were subsumed under the category "other relatives." Thus, they could not be identified and included in this category.

The next three variables describe the family characteristics of households that receive remittances. The first one is the marital status of the head, which is coded in the following way: married people or people living in consensual unions are grouped in one category, and the rest — people who are single, separated, or divorced — are grouped in another category. This variable is not significant for any of the countries. The next variable — family type — tries to pinpoint the stage of the family cycle of the household. For the purpose of this analysis, the family cycle is divided into four stages. The first comprises recently formed families without children. The second stage consists of families with children too young to be in the labor force. The third stage includes those families in which the older child has reached an age at which he/she can participate in the labor force. The last stage is one in which all the offspring have reached an age in which they can join the labor market. This variable was operationalized in the following way: families without children were considered in the formative stage of the family.[7] Children younger than 15 were considered too young to join the labor force (but that does not mean that they do not in fact join). Offspring above this age were considered old enough to join the labor force. This variable is significant in Jamaica, Guatemala, and Haiti. In the three cases, families in the later stage of the cycle receive remittances more than families in the previous stages. The last of the family variables is the number of children. This variable shows that larger families receive remittances more than other families, but this variable is significant only in the Dominican Republic and Guatemala.

Thus, the family variables suggest that families in a late stage of the family cycle or families with a large number of children have wider access to remittances as a mechanism for their subsistence. These results suggest that as the children of these families grow up, instead of looking for a job in their own city, they migrate and support their families with remittances. The results also suggest that larger families are more able to afford the economic or emotional costs of sending members abroad. The latter coincides with Funkhauser's (1992) findings for El Salvador. Thus the demographic profile of the households that receive remittances is one of female-headed, older households in a late stage of the family cycle, or households with a large number of children.

In order to find the main socioeconomic predictors of remittances, we ran a logistic regression on remittance-receiving households. Households that receive remittances are coded 1, and the rest are coded 0. The independent variables include the demographic variables used in the bivariate analysis in Table 3, in addition to years of education and income. The data on income suffer from a large number of missing values — close to 30% of the sample in the cases of Jamaica and Haiti. In order to solve this problem, we ran an ordinary least squares (OLS) regression on income, including as independent variables gender, age, education, number of children, and type of occupation. Using the coefficients obtained in this regression, we calculated the predicted income for the missing cases.[8]

Our claim is that during the 1980s, as a result of the deep economic crisis, migration, and hence remittances, encompassed all the urban social strata. Therefore, we expect to find that neither of the socioeconomic variables is significant in predicting remittances. The results of the logistic regression appear in Table 4 below and show that the relationship between socioeconomic status and access to remittances differ between the different countries.

Only the case of the Dominican Republic corresponds to our expectations. In this case, education is not significant, while income has a significant but negative effect on the likelihood of a household to receive remittances, suggesting that the poorer households are those that depend on remittances for their subsistence. These findings correspond with earlier reports of the wide socioeconomic scope of the Dominican migration (Guarnizo 1992). The cases of Jamaica and Haiti show a different pattern. In both cases education has a positive and significant effect on the likelihood of households to receive remittances, while income is not significant. These findings suggest that those households with higher human capital are more likely to receive remittances. These results are puzzling since migration from these two countries, as shown in Table 2, is very large. Finally, the case of Guatemala also contradicts our expectations. In this case, income has a significant positive effect, while education is insignificant. These findings may be related to the newness of Guatemalan urban migration.

The work of Massey and his collaborators (1987) suggest an explanation for the variations found in these results. They found that rural households receive larger amounts of remittances than urban households, a finding that they attribute to the different types of jobs that urban and rural migrants sought. Urban migrants found employment in cities where the cost of living is expensive, hence they could not save much to send home.

TABLE 3: Demographic Characteristics of Heads of Households that Receive Remittances by Country

	Jamaica		Haiti		Dominican Republic		Guatemala	
	Percent	N[a]	Percent	N	Percent	N	Percent	N
Gender								
Men	15.8	406	37.5	122	24.0	250	11.7	248
Women	23.5	378	42.0	188	31.4	153	19.1	152
		$p < .01$[b]						$p < .05$
		V .09[c]		n.s.[d]		n.s.		V .10
Age								
40 or younger	17.1	391	35.4	158	25.8	155	10.0	150
41-60	18.4	265	45.6	114	22.2	167	10.2	187
61 or older	35.1	131	48.1	27	38.3	81	38.1	63
		$p < .01$				$p < .03$		$p < .01$
		V .17		n.s.		V .13		V .28
Marital status[e]								
Married	17.8	454	42.4	132	25.5	259	16.9	124
Nonmarried	21.6	338	38.7	168	29.2	144	13.4	276
		n.s		n.s.		n.s.		n.s.
Type of family								
No children	20.0	90	43.3	60	31.8	22	0.0	16
All < 15	14.6	261	31.4	118	23.0	135	7.4	121
Older > 15	15.7	153	36.5	52	31.8	88	10.3	107
All > 15	25.8	283	55.9	68	26.6	158	24.2	153
		$p < .01$		$p < .01$		n.s.		n.s.
		V .12		V .19				
Number of children								
0 - 2	16.3	355	39.2	148	26.6	143	11.0	100
3 - 5	20.7	299	41.6	125	22.5	187	12.0	225
6 or more	24.8	133	40.0	25	38.4	73	26.4	72
		n.s.		n.s.		$p < .04$		$p < .01$
		n.s.		n.s.		V .12		V .16

[a] Totals differ due to different number of missing values.
[b] Significance of χ^2 test
[c] Strength of Cramer's V test of association. The closer the coefficient is to 1, the stronger the association.
[d] Not significant
[e] The category "married" includes both married people and couples living together. "Nonmarried" includes the rest (bachelors, separated or divorced people, and widows).

TABLE 4: Logistic Regression: Determinants of Households That Receive Remittances

	Jamaica	Haiti	Dominican Republic	Guatemala
Gender	-.58**	-.09	.42	-1.02**
Age	.02**	.02	.03*	.05**
Education	.07*	.10**	.05	.01
Ln income	.05	-.02	-.16*	.67*
Number of children	.05	-.08	.11*	.08
Marital status	-.11	.30	.12	.58
Type of Family				
No children	.24	-.82	1.25*	-5.61
Children < 15	-.13	-1.28	.72	-.34
Older > 15	-.39	-.66	.87*	-.52
Model χ^2	32.58**	30.13**	24.87**	42.94**
Degrees of freedom	9	9	9	9
Concordant pairs (%)	80.46	65.52	73.13	85.89
Number of cases	778	290	402	397

* $p < .05$ ** $p < .01$

In a similar vein, the wide access to remittances in the Dominican Republic may be explained by the presence of a Dominican ethnic economy in New York, which offers employment to Dominican migrants. This ethnic economy, as shown by Portes and Guarnizo (1991), is tightly related to the country of origin. Relatively large amounts of capital and goods flow between New York City and the island. The results in the cases of Haiti and Jamaica suggest that only migrants with higher levels of human capital are able to secure incomes that allow them to send money home. Migrants with lower levels of education may engage in recurrent migration that allows them to save some money and return home, as is the case of Jamaican cane cutters in Florida, or in activities that provide merely a subsistence income, as is the case with Haitian cane cutters in the Dominican Republic. The Guatemalan migration flows are newer. In the beginning of a migration flow, the migrants are those who can afford the trip and have the human capital to insert themselves into the receiving society. Social networks in the receiving countries that allow people from all social strata to migrate may not be present yet for Guatemalans.

Remittances and Household Income

In examining the importance remittances have on household income, we claim that in the context of the switch to an export-oriented model of development, based on the comparative advantage of cheap labor, remittances become a central component in the household efforts to subsist or improve its situation. Hence, we expect remittances to constitute an important part of total household income, and we expect households that receive remittances to have a higher total income than those that do not.

For the purpose of this analysis we constructed a variable, "total household income," composed of the sum of the different household incomes — the income of the head of the household, secondary income,[9] and remittances. Cases with missing values in their reported remittances or secondary income were coded as having a zero income in these accounts. This mode of calculation of the total household income is bound to increase the importance of the income of the head of the household vis-à-vis remittances or secondary incomes. This is so because we calculated income values for the missing cases of the former variable. This conservative way of calculating the importance of remittances increases the confidence in our results, which are presented in Table 5.

In Table 5, we should note the very low mean income of the heads of household in the four cities. Paradoxically, the Haitian sample has the highest mean monthly income, though even in this case the mean income is rather low.[10] In the four cases the results support our expectations concerning the importance that remittances have on household well-being. In the four cases, remittances are an important component of total household income; from about one fourth in Haiti to nearly 40% in the Dominican Republic. Moreover, the total income of households that receive remittances is considerably higher than the total income of households that do not receive remittances. These results demonstrate that in a context of deteriorating labor markets, those low-income households that have access to remittances can do considerably better than those that do not.

The results of a comparison of the mean income of the heads of the household in the four cities, correspond with our findings in the previous section, with the sole exception of Haiti. In the Dominican Republic, the mean income of the heads of household that do not receive remittances is higher than that of those households that receive remittances, showing again that in this country the poorest among the urban poor are those that depend on remittances. These results coincide with those of Massey and his collaborators (1987), who found that in the Mexican urban communities the households that relied on remittances were those with the lowest income and the worst labor market prospects.

In Guatemala and Haiti, the situation is reversed: the better off among the poor are those that receive remittances. Table 4 above has already shown that this was the case for Guatemala. In Haiti, education seems to be a better predictor of a household receiving remittances, but Table 5 shows that there are also differences in income. In the case of Jamaica, the mean incomes of both types of household are almost equal. *T* tests conducted on these means were significant in the cases of the Dominican Republic, Guatemala, and Haiti, and

TABLE 5: Means of Different Household Incomes by Remittances[a]

Jamaica

	Household Receives Remittances Yes	Percent	Household Receives Remittances No	Percent
Total income	254.9	100.0	165.9	100.0
Head's income	118.5	46.4	128.9	77.6
Remittances	84.9	33.3	0.0	0.0
Secondary income	51.8	20.3	37.0	22.4
Number of cases	154		638	

Haiti

	Household Receives Remittances Yes	Percent	Household Receives Remittances No	Percent
Total income	480.1	100.0	331.3	100.0
Head's income	310.1	64.6	263.9	79.6
Remittances	111.4	23.2	0.0	0.0
Secondary income	58.5	12.1	67.3	20.3
Number of cases	121		179	

[a] Monthly income in 1991 US$

not significant in the case of Jamaica, showing that for the latter case, the mean income of the two types of household is essentially equal.

Remittances and Labor Force Participation

In this section, I examine the effects of remittances on the labor force participation of the head and other members of the household. Our claim is that in the context of high rates of unemployment and underemployment, remittances allow households to send fewer people into the labor force, thereby reducing unemployment and underemployment pressures that are very high in these countries. Remittances also allow people to avoid engaging in the least desirable jobs; these are the trends in El Salvador (Funkhauser 1992).

Table 6 reveals the proportion of heads of household in different occupational categories that receive remittances. The occupational categories used in this article are the following: self-employed, owners, formal workers, informal workers, and people who do not have a regular job. The table shows that households in every occupational category receive remittances. In three of the

TABLE 5: Means of Different Household Incomes by Remittances[a] (Continued)

Dominican Republic

	Household Receives Remittances		Household Receives Remittances	
	Yes	Percent	No	Percent
Total income	281.0	100.0	219.3	100.0
Head's income	135.9	48.4	170.1	77.5
Remittances	107.3	38.2	0.0	0.0
Secondary income	37.8	13.4	49.2	22.5
Number of cases	108		295	

Guatemala

	Household Receives Remittances		Household Receives Remittances	
	Yes	Percent	No	Percent
Total income	345.2	100.0	163.0	100.0
Head's income	167.2	48.4	130.8	80.2
Remittances	101.5	29.4	0.0	0.0
Secondary income	76.4	22.1	32.2	19.7
Number of cases	58		342	

[a] Monthly income in 1991 US$

countries, however, the largest group is in the "not working" category.[11] These differences are statistically significant in the cases of Santo Domingo and Kingston. These results are consistent with our claim. In Haiti, the exceptional case, owners are the group with the largest access to remittances. Nevertheless, there is also a sizable group of remittance-receiving households in the "not working" category in Haiti.

Table 7 tests the hypothesis concerning the effects of remittances on the employment of the head of household. The table presents the results of a logistic regression on the labor force participation of the household heads. To construct this dependent variable, we dichotomized the occupations variable. People in the category "not working" were coded 1, and the rest were coded 0. In this analysis we used a dichotomous variable to test the effects of remittances; households that receive remittances were coded 1, the rest were coded 0. Additional independent variables include gender, age, education, the number of relatives abroad,[12] the presence of a secondary income, and the number of people in the household who work in addition to the head of the household.

TABLE 6: Occupation of Head of Household by Remittances

	Jamaica	Haiti	Dominican Republic	Guatemala
	Percent Receive Remittances	Percent Receive Remittances	Percent Receive Remittances	Percent Receive Remittances
Occupation[a]				
Self-employed	14.6	40.4	22.7	11.7
Owner	19.4	60.0	16.1	12.5
Informal	14.4	32.7	17.5	10.2
Formal	16.3	17.6	30.0	12.8
Not working	27.2	43.4	37.4	22.4
Number of cases	152	121	107	58
χ^2 significance	$p < .01$	n.s.	$p < .02$	n.s.
Cramer's V	.14		.17	

[a] Self-employed are people who conduct their own business without employing anyone else. Owners have their own business and employ others. Informal workers are wage workers not covered by social security. Formal workers are wage workers covered by social security. Not working includes the following categories: unemployed, living from a rent, living from a pension, housewives, students, and other.

The results of the logistic regression show that gender, age, and remittances are the three main predictors of nonparticipation in the labor force. Women and older people, as could be expected, are more likely to be in this situation. In Kingston, Port-au-Prince, and Santo Domingo, remittances have a significant positive effect on nonparticipation. In Guatemala, the effect of remittances is in the same direction, but is not statistically significant. In a cross-sectional study such as this one, it is difficult to infer causality, but the results show a clear negative relationship between receipt of remittances and the labor force participation of the head of household.

A proper interpretation of these findings needs to take into account the labor market context in which this study took place. It is difficult to overestimate the social costs of the changes that have occurred during the last decade. The burdens that the economic crisis and subsequent policy changes have imposed on households are numerous, and have been aptly described elsewhere (Beneria & Feldman 1992). Here we provide some statistics that highlight the labor market situation in these cities. In 1989, open unemployment in Guatemala City was 6.2%, while 33.0% of the labor force was engaged in informal activities. In the same year in Kingston, 11.4% of the men in the work force and 21.8% of the women were unemployed, while 26.0% of the labor force was in the informal sector. In 1991, 26.8% of the urban labor force was unemployed in the

TABLE 7: Logistic Regression: Predictors of Heads of Household Not Holding a Regular Job

	Jamaica		Haiti		Dominican Republic		Guatemala	
	B	Δp[a]	B	Δp	B	Δp	B	Δp
Gender	-1.40**	-.38	-.99**	-.22	-1.08**	-.08	-2.15**	.16
Age	.04**	.01	.01		.07**	.01	.08**	.01
Education	-.07*	-.02	-.06*	-.01	-.01		.02	
Number of people in the household who work[b]	-.15	-.03	-.29	-.07	-.03		.16	
Number of relatives abroad	-.10		-.07	-.02	-.24*	-.01	-.12	
Receives remittances	.48*	.09	.59*	.14	.73*	.02	.21	
Has secondary income	.09		-1.01**	-.22	.67*	.02	.40	
Model χ^2	162.41**		50.36**		106.02**		133.19**	
Degrees of freedom	7		7		7		7	
Concordant pairs (%)	73.78		68.04		80.35		77.00	
Number of cases	778		290		402		397	

[a] Indicates the actual probability of the event associated with a unit change in the independent variable. Calculated according to the formula: $\exp(L_1)/(1 + \exp(L_1) - \exp(L_0)/(1 + \exp(L_0)$. L_0 is the logit before the unit change in X_j, and L_1 is the logit after the unit change in X_j (see Petersen 1985).
[b] Does not include the head of the household.

* $p < .05$ *** $p < .01$

cities of the Dominican Republic. Another 25.2% was self-employed. The numbers for Haiti are scarce and unreliable, but it is well known that most of the Haitian economy is informal (Portes et al. 1994). In addition, it is helpful to remember that the means of total monthly income for the four cities are as follow: US$235.9 in Santo Domingo, US$189.1 in Guatemala City, US$183.3 in Kingston, and US$391.0 in Port-au-Prince.

In the labor markets of the cities in the periphery, the notion of unemployment is not equivalent to that of the cities in the core. In places where there is no safety net for unemployed people, most of the people who claim not to work are engaged in some sort of gainful activity, encompassing such diverse ventures as sewing and cooking for the neighbors, organizing rotating saving groups, working day-labor on occasional construction or repair jobs, or selling clandestine lottery tickets (Itzigsohn 1994).

Our results are consistent with the hypothesis that in a labor market situation where the number of well-paid stable jobs is being reduced, remittan-

ces allow heads of household to avoid unwanted jobs. This in turn frees them to engage in all sorts of casual labor that may provide a similar or better income, albeit less secure, than low-paying jobs in the public sector or export assembly manufacturing. In addition, these activities permit more flexible work schedules and sometimes provide the gratification of working independently. This in turn relieves unemployment pressures. The exceptional case of Guatemala, where remittances do not have a significant effect, supports this interpretation. As we saw above, people with a higher income in Guatemala are those who have access to remittances. They have a better labor market position, and so they are not led out of the labor force by remittances. Finally, we are interested in testing the effects of remittances on the labor force participation of additional members of the household. Our claim is, again, that remittances reduce the pressures on the household to insert additional members into the labor market. To test this claim we ran an OLS regression on the number of people in the household who participate in the labor force in addition to the head. The regression includes the individual and household variables included thus far, plus two new variables: the natural logarithm of the income of the head, and a series of dummy variables representing the different types of occupation.[13]

Table 8 presents the standardized results of the regression. In Santo Domingo and Guatemala, the results conform to our expectations. Remittances have a negative and statistically significant effect on the number of people who contribute to the income of the household. In Kingston the effect is in the expected direction, but it is not statistically significant. Only in Port-au-Prince does this variable have a positive coefficient, but not a statistically significant one.

Port-au-Prince is the exceptional case. As shown in Table 6, a large proportion of owners receive remittances in that city. This fact suggests that remittances may be used as a source of capital accumulation by the more financially secure households.[14] These, however, are informal owners who tend to employ family members in their business as a form of primitive capital accumulation — hence the positive sign of the effect of remittances on the labor force participation of household members.

Overall, these results suggest that in most cases, households that receive remittances need to send fewer additional members abroad than those that do not receive remittances. Again, this is important because of the same reasons mentioned above regarding the heads of household. In addition this may allow younger members of the household to stay in school and complete their studies.

In order to test this hypothesis we ran an OLS regression on the ratio of school-age offspring who are in school to the total number of school-age offspring in the household.[15] The independent variables include the independent and the dependent variables used in the previous analysis. The results are not presented because the effects of remittances are not significant in any of the cases.[16] Nevertheless, it is interesting to report the direction of the effects. Remittances have a positive impact on children's remaining in school in Santo Domingo and Kingston, and a negative effect in Guatemala and Port-au-Prince. The negative effect in Port-au-Prince is not surprising given the positive effect of remittances on the number of members of the household in the labor force.

TABLE 8: OLS Regression: Predictors of the Number of People in the Household Who Work

	Standardized Coefficients			
	Jamaica	Haiti	Dominican Republic	Guatemala
Gender	-.08*	.08	-.12*	-.16**
Age	-.05	-.17**	.09	.14**
Education	-.09*	-.03	-.03	-.08
Ln income	.01	.08	.01	.06
Number of children	.23**	-.11	.19**	.24**
Marital status	.10**	-.01	.01	.07
Self-employed	-.02	.26*	-.09	-.07
Informal worker	-.06	.14	-.12*	-.06
Owner	-.01	.19*	.01	.06
Do not work	-.07	.05	-.03	.04
Receive remittances	-.01	.07	-.11*	-.17**
R^2	.08	.13	.10	.17
Number of cases	778	290	402	397

* $p < .05$ ** $p < .01$

The negative effect in Guatemala is more difficult to explain. Santo Domingo and Kingston, however, are the two cities where the lower-income groups have a wider access to remittances. These results, while far from conclusive, provide some support to the hypothesis that remittances help the weakest groups to accumulate human capital.

Conclusions

At the beginning of this article we observed the increasing importance of remittances to the economies of the countries of the Caribbean Basin. Our findings show the social effects of that economic phenomenon. In each of the cities of the study, a large number of households depend on remittances for subsistence or mobility. Remittances are not only an important macroeconomic factor, helping to ease the pressures on the current account, but also they affect the lives of a large number of middle-low-income and low-income people. We also found that, against our expectations, only in the Dominican Republic do remittances reach equally every socioeconomic stratum. In the other three countries, groups with higher income and/or human capital have wider access to remittances. We explain these findings by relating the scope of the availabili-

ty of remittances to the employment opportunities of the migrants rather than to the extent of migration.

As we expected, remittances are a very important component of household income in these social sectors. Households that receive remittances have a higher average income than households that do not receive remittances in the four cities. Remittances thus, are central to the well-being of middle-low and low-income households. We also found that in Santo Domingo, Kingston, and Port-au Prince, remittances allow people to stay out of the labor force. Also in Santo Domingo, Kingston, and Guatemala City, remittances allow households to send fewer people into the labor force. We infer that these effects alleviate unemployment pressures in labor markets characterized by high rates of unemployment and underemployment. Moreover, remittances allow the people in the weakest labor market positions to avoid the worst menial jobs and engage in more flexible forms of employment.

The patterns we found are not similar in every city. We have attempted to explain the divergences. Overall, however, our findings provide support to our claim that in the labor market context of the beginning of the 1990s, remittances constituted a central strategy for the subsistence of low-income households in the Caribbean Basin. Sending members abroad in order to receive remittances is a rational choice of people confronting very difficult circumstances. Those labor market conditions have not changed since then, neither are they likely to change as long as these countries compete for scarce capital through the cheapening of their labor costs.

The main implication of these findings is that an increasing number of households, willingly or reluctantly, are going to send members abroad so that the latter can send back remittances. The most likely destination has been, and will continue to be, the U.S. As large Caribbean migrant communities grow in this country, they provide social networks that facilitate a continuing flow of migrants, in spite of the increasingly hostile attitude of large segments of the U.S. public.

By the end of the 1970s, the import substitution industrialization model implemented in the area since the 1950s was indeed exhausted, and a new form of entry into the world economy was needed. The new direction adopted by the countries of the region was a restructuring of their economies with a new emphasis on export-oriented activities, the bulk of them directed toward the U.S. The Caribbean Basin Initiative (CBI) played an important role in easing the entry of Caribbean exports into American markets. This new mode of entry into the world economy may, in some cases, alleviate the burden on the current account of peripheral countries. It does not, however, increase employment or income. As the societies and the economies of the region become increasingly integrated, it is in the interest of the Caribbean countries, as well as the U.S. — the main recipient of the migration flows — to make a joint effort to find better means of economic integration, to find a development strategy based on the development of the human capital of the region and on dynamic, rather than static, comparative advantages.

Notes

1. San Jose, the capital of Costa Rica, was also part of the study, but due to the low importance of migration from Costa Rica, it is not included in this article.

2. In this work, the term *maquila* refers to firms that, although not located in export zones, receive the benefits granted to the latter in terms of facilities for import and export assembly manufacturing products. This is important to note because sometimes the term *maquila* is used to refer to any firm that does subcontracting work for others.

3. This decline is probably the result of the growth of exports as a result of the export-oriented development policies.

4. In Jamaica, where updated statistical information is available, the selection of neighborhoods was made according to measurable criteria. In the other three countries the selection was conducted according to whatever statistical information was available, in addition to the knowledge of the city of the research teams, which have been conducting research in their countries for long years. Post-hoc tests of the characteristics of the sample in the different neighborhoods confirmed that the choices correspond with the researchers' assumptions.

5. The respondents had to be 30 years of age or older. This is so because the survey attempted to capture the changes in the economic position and the perceptions of the respondents as a result of the crisis of the 1980s.

6. The original questionnaire was written in Spanish, and translated by the research teams into English (Jamaica) and Haitian Creole.

7. Single people without children are included also in this category. The reason for this is their small number in the sample.

8. I prefer this method to others, such as assigning the mean value to the missing cases, because in this case the predicted income is related to the characteristics of the individuals and the sample. In order to check that no major distortion of the data was caused by this method of calculation, I ran a series of t tests comparing mean incomes of different groups. For this I used the new variable that includes all the cases and the original variable, which included only the cases that reported income. I compared the means for gender, occupation, and whether the household receives remittances. The results of these tests show that in all cases the means of the new variable are slightly higher than the original ones, but they reproduce the same patterns for the different groups as the original ones. These results are available upon request from the author, and they suggest that the new income variable can be used with confidence.

9. Secondary income refers to contributions to the household income earned locally by people other than the head of the household.

10. We think this result is paradoxical since Haiti is the poorest country of the four included in this study. This could be the result of the presence of outlier cases in the sample, or of a high exchange rate for the local currency at the time of the survey (incomes were reported in local currencies and transformed into US$ using the official exchange rate of the time of the survey).

11. The category "not working" is a broad one. It includes people that are actually searching for a job and people that are outside of the economically active population (e.g., housewives, people living off a pension or other sources of income not related to current employment).

12. The number of relatives abroad is a continuous variable, ranging from zero for households without relatives abroad, to three for households with three or more relatives abroad. The limit of three is due to the fact that the questionnaire did not ask about the total number of relatives abroad, but about the relationship and whereabouts of the closest three relatives of the respondent who reside abroad.

13. The effects of the dummy variables are not absolute effects, but should be compared to a reference category that is omitted from the regression. In this case, the omitted reference category is formal workers.

14. This interpretation coincides with the findings of Portes and Guarnizo (1991) for the Dominican Republic and Massey et al. (1987) for Mexico.

15. School-age offspring are those older than 5 and younger than 19.

16. These results are available upon request.

References

Appleyard, Reginald. 1989. "Migration and Development: Myths and Reality." *International Migration Review* 23:486-99.

Beneria, Lourdes. 1992. "The Mexican Debt Crisis: Restructuring the Economy and the Household." Pp. 83-104 in *Unequal Burden: Economic Crisis, Persistent Poverty, and Women's Work*, edited by Lourdes Beneria and Shelley Feldman. Westview Press.

Beneria, Lourdes, and Shelley Feldman. 1992. *Unequal Burden: Economic Crisis, Persistent Poverty, and Women's Work*. Westview Press.

Birks, J.S., and C.A. Sinclair. 1980. *International Migration and Development in the Arab Region*. Geneva: ILO.

Bray, David. 1987. "Agricultura de Exportacion, Formacion de Clases y Mano de Obra Excedente: El Caso de la Migracion Interna e Internacional en la Republica Dominicana." Pp. 92-108 in *La Inmigracion Dominicana en los Estados Unidos*, edited by J. del Castillo and C. Mitchel. Santo Domingo: Editorial CENAPEC.

Funkhauser, Edward. 1992. "Mass Emigration, Remittances, and Economic Adjustment: The Case of El Salvador in the 1980s." Pp. 135-75 in *Immigration and the Work Force: Economic Consequences for the United States and Source Areas*, edited by George Borjas and Richard Freeman. University of Chicago Press.

Ghai, Dharam, and Cinthya Hewitt de Alcantara. 1990. "The Crisis of the 1980s in Sub-Saharan Africa, Latin America and the Caribbean: Economic Impact, Social Change and Political Implications." *Development and Change* 21:389-426.

Grasmuck, Sherri, and Patricia R. Pessar. 1991. *Between Two Islands: Dominican International Migration*. University of California Press.

Guarnizo, Luis E. 1992. *One Country in Two: Dominican-Owned Firms in New York and the Dominican Republic*. Ph.D. diss., Department of Sociology, Johns Hopkins University.

Hamilton, Nora, and Norma Stoltz Chinchilla. "Central American Migration: A Framework for Analysis." *Latin American Research Review* 26:75-110.

Kaimowitz, David. 1990. "The Political Economy of Central America: Foreign Aid and Labor Remittances." *Development and Change* 21:637-54.

Keely, Charles B., and Bao Nga Tran. 1989. "Remittances from Labor Migration: Evaluations, Performance and Implications". *International Migration Review* 21:500-25.

Kritz, Mary M., and Charles B. Keely. 1981. "Introduction." Pp. XIII-XXXI in *Global Trends in Migration*, edited by Mary M. Kritz, Charles B. Keely and Silvano M. Tomasi. Staten Island: Center for Migration Studies.

Lopez, Jose Roberto, and Mitchell A. Seligson. 1991. "Small Business Development in El Salvador: The Impact of Remittances." Pp. 175-206 in *Migration, Remittances, and Small Business Development. Mexico and Caribbean Basin Countries*, edited by S. Diaz-Briquets and Sidney Weintraub. Westview Press.

Lundahl, Mats. 1985. "International Migration, Remittances and Real Incomes: Effects on the Source Country." *Scandinavian Journal of Economics* 87:647-57.

Martin, Philip. 1991. "Labor Migration and Economic Development." Pp. 241-58 in *Determinants of Emigration from Mexico, Central America, and the Caribbean*, edited by S. Diaz-Briquets and Sidney Weintraub. Westview Press.

Massey, Douglas, Rafael Alarcon, Jorge Durand, and Humberto Gonzalez. 1987. *Return to Aztlan: The Social Process of International Migration from Western Mexico*. University of California Press.

Petersen, Trond. 1985. "A Comment of Presenting Results from Logit and Probit Models." *American Sociological Review* 50:130-31.

Portes, Alejandro, and Luis Guarnizo. 1991. "Tropical Capitalists: U.S. Bound Immigration and Small Enterprise Development in the Dominican Republic." Pp. 101-32 in *Migration, Remittances, and Small Business Development in Mexico and the Caribbean Basin Countries*, edited by S. Diaz-Briquets and Sidney Weintraub. Westview Press.

Portes, Alejandro, and Ramon Grosfoguel. 1994. "Caribbean Diasporas: Migration and Ethnic Communities." *Annals of the American Academy of Political and Social Sciences 533* (May):48-69.

Portes, Alejandro, Jose Itzigsohn, and Carlos Dore. 1994. "Urbanization in the Caribbean Basin: Social Change during the Years of the Crisis." *Latin American Research Review:* 3-37

Programa Regional de Empleo para America Latina y el Caribe. PREALC. 1987. *Ajuste y Deuda Social. Un Enfoque Estructural*. Santiago de Chile: Organizacion Internacional del Trabajo.

Przeworski, Adam, and Henry Teune. 1970. *The Logic of Comparative Social Inquiry*. Wiley.

Roberts, Brian. 1989. "Employment Structure, Life Cycle, and Life Chances: Formal and Informal Sectors in Guadalajara." Pp. 41-59 in *The Informal Economy: Studies in Advanced and Less Developed Countries*, edited by Alejandro Portes, Manuel Castells, and Laureen Benton. Johns Hopkins University Press.

Safa, Helen I., and Peggy Antrobus. 1992. "Women and the Economic Crisis in the Caribbean." Pp. 49-82 in *Unequal Burden: Economic Crisis, Persistent Poverty, and Women's Work*, edited by Lourdes Beneria and Shelley Feldman. Westview Press.

Sassen, Saskia. 1988. *The Mobility of Labor and Capital: A Study in International Investment and Labor Flow*. Cambridge University Press.

Selby, Henry A., and Arthur D. Murphy. 1982. *The Mexican Urban Household and the Decision to Migrate to the United States*. ISHI Occasional Papers in Social Change #4. Philadelphia: Institute for the Study of Human Issues.

Stahl, Charles, and Fred Arnold. 1986. "Overseas Workers' Remittances in Asian Development." *International Migration Review* 20:899-925.

Straubhaar, Thomas. 1992. "The Impact of International Labor Migration for Turkey." Pp. 79-131 in *Migration and Economic Development*, edited by K.F. Zimmermann. Berlin: Springer-Verlag.

Tardanico, Richard. 1993. "Dimensions of Structural Adjustment: Gender and Age in the Costa Rican Labour Market." *Development and Change* 24:511-39.

World Bank. 1993. *World Tables 1993*. Johns Hopkins University Press.

Why Do Daughters Leave School in Southern Africa? Family Economy and Mothers' Commitments*

BRUCE FULLER, *Harvard University*
JUDITH D. SINGER, *Harvard University*
MARGARET KEILEY, *Harvard University*

Abstract

By focusing on the family economy and its capacity to mediate broad economic developments, sociologists are beginning to move beneath macrostructural forces to better explain parental demand for schooling and children's attainment. This materialist model focuses on the explanatory power of contextual labor demand and resources internal to the family. Parents' social preferences and commitments, antecedent to "choosing" between work or school for their children, are presumed to covary with economic factors. In contrast, research on family practices within impoverished settings reveals that parents' social commitments linked to child development or schooling can vary independently of the family's economic circumstances. Applying these alternative theories to family behavior in southern African, we assess the relative influence of mothers' economic demands and social commitments on their daughter's probability of staying in school. We find that the risk of daughters leaving school is more strongly influenced by mothers' social commitments than by household economics. Maternal influences do interact with selected family-economy indicators and are conditioned by the level of discretionary time afforded to daughters.

Children around the world, and within particular societies, leave school at widely varying ages. Efforts to identify the underlying causes of why, and when, children are likely to exit school have greatly influenced two competing

* *David Cownie and Elizabeth Blake assisted enormously in the design and implementation of this study. Household interviews were conducted with essential help from Nakiso Kubanji, Matilda Mangole, and Boineelo Mokgothu. Encouragement and support from Barbara Belding, Stephanie Funk, Lynellyn Long, Chloe O'Gara, and May Rihani are greatly appreciated. Conversations with, and related research of, Susan Holloway, Anjimile Kapakasa, Karin Hyde, Robert and Sarah LeVine have influenced our work. Direct correspondence to Bruce Fuller, 6 Appian Way, Graduate School of Education, Harvard University, Cambridge, MA 02138.*

models of social action and attainment. Individual-level *rational choice* theorists posit that parents actively construct strategies for maximizing the family's economic returns, investing in their children's human capital within the household's budget constraint (Becker 1976; Coleman 1986; Schultz 1961). In contrast, *macrostructural* proponents argue that national or local opportunity structures — reproduced through dominant forms of status attainment, labor demand, social-class membership, or state-constructed supply of schooling — determine how far children go in school (Garnier, Hage & Fuller 1989; Hauser & Daymont 1977; Ogbu 1978; Rubinson 1986).

Recently the *family economy* has received attention from social historians and sociologists — as an intermediary organization that actively accommodates novel labor demands arising within industrializing economies (Hareven 1982; Horan & Hargis 1991; Robinson 1993; Walters & O'Connell 1988). This new conception of family economy — and its capacity to explain change in the allocation of children's time between school and work — draws on the household economist's basic model: parents intentionally advance their child's interests, mediating labor demands and seeking income benefits found within a modernizing environment. This model, however, also draws on institution-level work, emphasizing that the range of strategies available to the family is bounded by dominant forms of production, labor demand, and political organization that comprise the opportunity structure (Carnoy & Levin 1985; Walters & James 1992).

Research on school attainment within impoverished settings takes a different tack, emphasizing that family strategies and parents *social commitments* are further constrained by forces that vary among households within a particular social class. Among low-income parents, for instance, time spent supervising children's schoolwork in the U.S. (Astone & McLanahan 1991) and parental endorsement of the school itself, cross-culturally, can vary significantly (Goodnow & Collins 1990). This alternative framework suggests that parents' social commitments are influenced by economic demands; yet their acquired beliefs may operate somewhat independently. When the family's institutional commitments do vary independently of economic position, neither deterministic macrostructural nor intention-filled family-economy explanations of attainment are complete.

Our study of daughters' educational attainment in southern Africa advances this debate over how to explain family demand for schooling by (1) assessing the independent effects of economic demand *and* social commitments held by parents, (2) isolating parental beliefs that are specific to schooling and literacy, rather than seeing parental values and preferences as psychological phenomena uncoupled from particular institutions, and (3) focusing on how mothers' economic resources, gender-specific labor demands, and social beliefs help to explain their daughter's attainment. We first discuss how the family-economy framework can be broadened. We then apply this more complete, and gender-specific, model to the commercializing society of Botswana.

Family Economy and Parental Commitments

MATERIALIST EXPLANATIONS OF FAMILY DEMAND FOR SCHOOLING

The family-economy model, as it evolves to explain school attainment levels, has made four distinct conceptual advances: (1) in contrast to macrostructural explanations of demand for schooling, family-level "choices" and children's resulting enrollment (or work) behavior are directly observed; (2) competing demands placed on the family are observed historically, looking out into a heterogenous environment from parents' own viewpoint, not simply inferring uniform family responses to aggregate labor-demand shifts or action by the state (Walters & James 1992; Walters & O'Connell 1988); (3) departing from the neoclassical household-economics model, parents are viewed as enmeshed in economic interdependencies *and* social obligations, rather than assuming that parental decisions about their child's schooling are made individually for each child in a normative vacuum (Hareven 1982; Tilly & Scott 1978); and (4) the influence of gender-related attributes of the family on attainment has been introduced, including effects of mother's income (e.g., from boarders) and gender composition of the household (Horan & Hargis 1991; Robinson 1993).[1]

The sociological version of family-economy does converge with the earlier household-economics model. Both assume that the structure of production and work is the principal determinant of schooling demand. Walters and O'Connell (1988) point out that human capital theorists have erred by focusing on *long-run* returns to schooling (presumably calculated by parents), as opposed to *immediate* production or income needs faced by the family to which children can contribute. But the argument remains principally a materialist one: parents' demand for schooling flows from short-term income and labor needs, involving their children's contribution to subsistence farm production or to the family's cash earnings. Both versions have recently introduced the supply and quality of the *school institution* as a factor in the formulation of family strategies, linking to actions by the state (Knight & Sabot 1990; Walters 1992). Thus, within modernizing environments the family is seen as mediating economic *and* contested institutional norms about how children should normatively spend time between work and school. The macrostructure is variegated both in terms of diverse local economic conditions and in the range of possible social commitments that may be extracted by parents. It is, therefore, not surprising that individual households vary in how they negotiate these environs (Hernes 1976).

THE FAMILY INSTITUTION'S COMMITMENTS: BROADENING THE FAMILY-ECONOMY MODEL

Our study applies a broader conception of the family-economy model to the issue of school attainment and parental decision making in contemporary Botswana. We briefly outline how the basic model can be extended to include material *and* social factors operating within the household which may influence the child's propensity to stay in school. First, early research on school attainment pointed to the role of the child's social-class membership in predicting how far she or he went in school (Sewell & Hauser 1975). Only recently have survey researchers attempted to uncover *why* social class matters and *how* family

strategies or specific parenting practices differentially influence children's commitment to and performance in school (Astone & McLanahan 1991; Baker & Stevenson 1986). Comparative research reveals that parents' endorsement of the school institution and their enabling behaviors (e.g., time spent helping children with homework) vary significantly among families *within* social class groups (Goodnow & Collins 1990), and parental beliefs directly linked to schooling more consistently predict children's attainment, compared to the earlier focus on parents' general values regarding modernity or individualism (for review, Hess & Holloway 1984).[2]

Second, the family-economy framework emphasizes material demands faced by, and resources available to, parents. The surrounding structure of (subsistence or industrial) production and labor places demands on the family, depending on their position within the occupational structure (Horan & Hargis 1991). The assumption is that work demands placed on children by parents significantly cut into time available for schooling (the influence of opportunity costs and the budget constraint within the neoclassical model, Becker 1976). However, parents' commitments (e.g., endorsement of the school's utility) represent socially constructed "resources" that also may influence school attainment, independent of wealth or macro labor demand. Family structures hosting these social resources are reproduced not only through economic processes but also through institutionalized cultural scripts and commitments (Angel & Tienda 1982; LeVine 1987). While child labor demands can be significant within impoverished communities (historically) or in peasant societies (currently), *discretionary time* made available to children varies widely across local settings.

Family-economy theorists have yet to confront the issue of how the social meaning and behavioral correlates of class position may be conditioned by local culture. The class position of fathers in contemporary Western societies may be related to individual values and child-rearing beliefs (Kohn et al. 1990). But it remains unclear whether typical indicators of class membership (wage-sector occupation, cash income) provide useful predictors of children's school attainment in non-Western societies.[3] How mothers and daughters spend time in the household, and the amount of discretionary time made available for school, may be reinforced by gender-specific cultural scripts that characterize the family institution; but this process may be only loosely coupled with the father's position in the wage economy.

Research on school attainment in developing countries similarly emphasizes household economy factors, especially when focusing on gender inequality. Studies of Third World households show that mothers and daughters may spend twice the time on labor tasks each week, relative to male members (Dixon-Mueller 1985; Hill & King 1993). Hyde's (1993) empirical review, however, pushes further to show that social and institutional factors also influence female attainment in certain settings: local norms regarding marriage and childbearing, religious beliefs linked to literacy, parents' selection of which children are encouraged to stay in school, and quality of the local school (also, for Kenya and Tanzania, see Knight & Sabot 1990; for Ghana, Glewwe 1991).

This leads to the third area: the family-economy framework can be pushed forward to carefully assess gender dynamics within the household. Both

economic and social demands placed on mothers differ from those faced by fathers. And parents' definition of, and strategies for advancing, daughters' status may differ relative to sons'. Where wage-labor demand systematically varies for female versus male graduates, parental decision-making related to school attainment predictably varies based on gender (Horan & Hargis 1991; Rosenzweig & Schultz 1982). In many developing countries, the "social opportunity cost" of staying in school may be high, where postponing marriage beyond age 15 or 16 can have sharply negative consequences for the daughter's and parents' social status. Mothers also may mediate labor and social demands facing their daughters; for instance, initial Third World evidence suggests that girls' school attainment is higher when the mother controls household spending, often within father-absent households (Schultz 1990).[4]

Our empirical study of daughters' school attainment in southern Africa first applies a basic *family-economy* model to contemporary households (equation 1). Then, we study a broader model that focuses on *gender-specific elements* of the family economy (equation 2). Finally, we assess a *family-institution* model which adds social commitments held by the mother which may further explain daughters' school attainment. We turn next to specific conditions found within rural Botswana villages, and then technically specify the three models.

Explaining Daughters' Attainment in Southern Africa

BOTSWANA CONTEXT

Botswana lies immediately to the north of the Republic of South Africa. About 85% of its 1.5 million inhabitants are from the Tswana tribe; a greater number of Tswana people live within South Africa and neighboring Namibia. Botswana has a strong cash economy, relative to other sub-Saharan African countries. Gross national product per capita equalled US$1,050 in 1987, but income inequalities are enormous. Botswana's rapid commercialization stems from rich diamond mines and a sizable beef-export sector. The proximity of wage jobs in South Africa has historically influenced labor migration of males, and thus family structure (Brown 1983). Two-thirds of the labor force still works outside the wage sector, primarily in livestock herding, farming, and the informal cash economy. Given that Botswana's rapid development disproportionately affects urban elites, economic and social practices remain traditional in many rural households. Despite a decline in the fertility rate, for instance, it remains high, with population still growing at 3.4% annually, compared to 2.4% in South Africa (Colclough & McCarthy 1980; World Bank 1991).

Females in Tswana culture have historically attended to domestic obligations: bearing and supervising children, preparing meals, and maintaining the family compound. Women quite frequently brew and sell beer to earn cash, especially in father-absent households (Alverson 1978). Adult and adolescent males have traditionally migrated to towns for cash-earning jobs, worked at the family's cattle post (grazing lands where herds are raised, an area that may be just outside the village or a day's walk), or tended the household's land plot (with female members). Time spent working the land is limited, however, given Botswana's arid climate. This allows children considerable discretionary time, a

crucial condition 'under which parents allocate time between work and schooling.

Government, since independence (1966), has rapidly expanded schooling: primary enrollments have quintupled, now equalling 85% of the age cohort. The number of secondary schools has grown from six to 140. At this level gender inequities arise: girls begin to leave junior secondary school (grades 8-9), many due to pregnancy or marriage. By age 19, 49% of all females are either mothers or experiencing their first pregnancy. By grade 9, boys are outperforming girls in most subjects (Duncan 1989; Kahn 1988).

DATA AND MODELS

In-depth interviews were conducted with 122 mothers, each having a daughter of secondary-school age. Ethnographic work also was conducted in one village, the findings of which are not reported here. A standard interview protocol was used which included forced choice and open-ended questions. Interviews were conducted within the household compound and in the indigenous language (Setswana), each lasting between 105 and 180 minutes. Within each of four villages, we randomly selected mothers of 15 girls who were presently attending junior secondary school; neighboring households were then visited until another 15 mothers were found, each with a school-age daughter who had dropped out of school before completing junior secondary school.[5] Literacy tests also were administered to the mother and "target" daughter.[6]

General Hazard-Rate Model

We analyze the effect of family-level factors on the risk of daughters' leaving school, using discrete-time survival analysis (Singer & Willett 1993; Tuma & Hannan 1984). Let T represent a discrete random variable indicating the grade when a girl exits school. In discrete-time survival analysis, rather than focus on the distribution of T directly, we focus on the distribution of discrete-time hazard, h_i, defined as the conditional probability that a randomly selected girl will drop out of school in grade j, given that she did not drop out of school before grade j:

$$h_j = Pr[T = j \mid T \geq j]$$

Because event occurrence is inherently conditional (a girl can drop out for the first time only if she has not dropped out already), this set of conditional probabilities represents the fundamental parameters of the discrete-time survival process that we study. Details of the model are available from the authors.

Basic Family-Economy Model

To test the basic family-economy model at the household level, predictor variables related to labor demands experienced by family members were included, as well as indicators of the household's material wealth or poverty.[7] We express this simple relationship between family-economic factors and the risk of leaving school as:

$$logit(h_j) = f(G,L,Y) \tag{1}$$

where the risk of leaving school in grade j [h_j] is expressed as a logistic function of grade in school [G], labor demands emanating from the local economy and mediated by parents [L], and income and resources available to the family unit [Y]. Expressed at the household level, this is the basic function advanced in the family-economy literature as reviewed above.

Gender Specific Family-Economy Model

Our second model improves upon this conceptualization by including variation in gender-specific work demands for females, and labor demands relative to males. Also included are culturally relevant measures of family resources that indicate not only aggregate income but also subsistence-level production, livestock holdings, and consumption preferences (separating long-term household investment from the possession of less durable consumer goods). This more complete and locally situated model is specified:

$$logit(h_j) = f(G, Lf, Lf/Lm, Of, Om, Yp, Yc, Yi) \quad (2)$$

where the risk of leaving school is expressed as a logistic function of labor demands placed on daughters within the household [Lf], the number of labor tasks allocated to females minus tasks allocated to males [Lf/Lm], wage-sector employment of the mother [Of] and father, if present for most of the year [Om], and wealth measured by volume of subsistence production [Yp], possession of nondurable consumer goods [Yc], and level of investment in, and quality of, the household compound [Yi]. Descriptions of these indices appear in the Appendix. Model 2 does not depart theoretically from the standard family-economy model; it does directly observe female-specific labor demands and resources that mothers might influence, on which initial Third World studies of attainment have focused (e.g., Hill & King 1993).

Family-Institution Model

The social structure, beliefs, and school-related practices of parents, especially those exercised by mothers, may vary independently of economic variables to further explain daughters' school attainment:

$$logit(h_j) = f(G, HA, AM, Sa, Sed, Med, Fed, Flit, Fread, Sq, Su, Ffert) \quad (3)$$

where the risk of leaving school stems from the number of adults residing full-time in the compound, presumably providing stronger supervision of children [HA], whether the father has left the household permanently or is absent more than two months of the year [AM], age and education level of the target daughter's siblings and father's education level [Sa, Sed, Fed], mother's education, assessed literacy level, and reading practices (i.e., materials and frequency with which the mother reads) [Med, $Mlit$, $Mread$], the quality of the local secondary school and utility of keeping one's daughter in school, as perceived by the mother [Sq, Su], and the mother's fertility practices [$Mfert$], including use of birth control methods and fertility rate, as indicators of attitudes toward child bearing and more individualized socialization of offspring (LeVine 1987).[8]

This third model supplements standard family-economy indicators of social beliefs and practices that may vary independently of the household's economic

status, as shown in the comparative research outlined above. We append the term "institution" in labeling the model since the reproduction of these social beliefs and cultural scripts occurs within the family organization *and* may be externally disrupted by alternative ideologies and scripts introduced by modern institutions (i.e., schools, health clinics, and the media).

As the analysis proceeded, two sets of interaction terms were studied. Household-economy and maternal social factors may interact with the grade level in which the daughter is enrolled, especially since the hazard of dropping out rises at the end of primary school (grade 7) and junior secondary school (grade 9). We also discovered that hazard rates varied depending on whether the family resided in a southern or northern village. So, southern residence was interacted with principal economic and social predictors. For all three models nonlinear forms of the predictors (two-degree polynomial terms) also were assessed.

DESCRIPTIVE FINDINGS

Table 1 summarizes descriptive data, splitting households between those where the target daughter was attending the junior secondary school (*in-school*) versus those where the daughter had left school (*out-of-school*). Just 54% of mothers surveyed were married. Of these households with married couples, 45% of the fathers were no longer full-time residents of the compound: they were living in town, South Africa, or at the cattle-post for two months or more during the year. Mothers with out-of-school daughters were significantly older than mothers with in-school daughters, although no differences were found in the age of the target daughter or in her birth order. On average, four adults and six children lived in the household.

Family Economy

Less than one-third of all mothers were employed in a cash-earning job. Among senior males in the compound (not necessarily the husband), 34% earned cash at least on a part-time basis. Mothers reported performing six domestic tasks alone each day, on average; another four tasks were performed together by the mother and daughter. In contrast, the household's senior male was responsible for one and a half daily tasks and performed one additional task with a household member. The family's involvement with formal markets is highly variable. About one-third took livestock to market, mostly cattle, goats, and chickens. The brewing and sale of traditional beer was a common market activity. About 60% of the households owned an ox-pulled plow, indicating at least subsistence production of sorghum or maize. These latter indicators of *household production* were combined into an index of household production.

Two other measures of household wealth were constructed. An index of *consumer goods* present in the household was developed, including items shown in Table 1. For example, 75% of all families owned a working radio, 34% had a bicycle. The enumerator also made an assessment of *household quality and investment*. Culturally specific items included in this index: presence of a square ("modern") house in the compound in addition to traditional round huts (78% of households), whether the thatch on the roof was of high quality, cut by a

craftsperson in a close-cropped manner (68%), whether only a primitive latrine-pit was apparent (10%), and whether a water standpipe existed within the compound (24%). This index of housing quality proved to be the only economic indicator consistently related to the daughter's probability of leaving school.

Schooling and Mothers' Social Beliefs

Mothers (but not senior males) differed in their level of school attainment (those with daughters in-school, 5.4 years; out-of-school, 3.5 years). Mothers with in-school daughters also performed significantly better on our brief literacy assessment, as did their daughters. Furthermore, two beliefs linked to schooling and fertility behavior were consistently related to daughters' school attainment: how the mother perceived the quality of the local secondary school, and whether the mother used a birth control device. Seven individual scales were used to assess mothers' perception of school quality (see Appendix). Mothers with daughters in school had higher perceptions of quality of the local school than mothers with daughters out of school. Furthermore, reported use of birth control was indeed related to the number of children to whom the mother had given birth. Mothers who used a birth control device had fewer live births (mean equalled 6.1) compared to mothers not using birth control (6.5).

HAZARD-MODEL FINDINGS: WHAT HOUSEHOLD FACTORS EXPLAIN DROP-OUT RATES?

The study of school attainment has been hampered by two analytic problems: variability in key outcomes — whether the child is still attending school and their years of attainment — is heavily dependent upon when the survey is conducted or the time points chosen for the analysis. Right censoring of cases often occurs for many individuals in the sample: the student may drop out the year after the survey is conducted, but this datum is never known. Nor can cross-sectional regression methods (or historical time-series models of school attainment) factor in variability associated with *when* children leave school. By applying hazard or event-history techniques, we can more adequately model the probabilities of *whether* and *when* girls leave school, based on cross-sectional survey data.[9]

Figure 1 plots the baseline hazard rate of leaving school by grade level as a percent of girls still in school relative to the previous year. For our sample, all but one girl stayed in school through grade 6; then 9% of the remaining girls dropped out prior to completing grade 7; and 26% of the remaining girls left school prior to grade 8 (the first year of junior secondary school). Institutional demarcations, particularly the transition from primary to secondary school, make a big difference. Among remaining female students, another 25% dropped out before or after completing grade 9. Substantive predictors are then added to this baseline function; improvements in fit can be evaluated by examining decrements to chi-square statistics.

TABLE 1: Characteristics of Botswana Households, Mothers, and Daughters

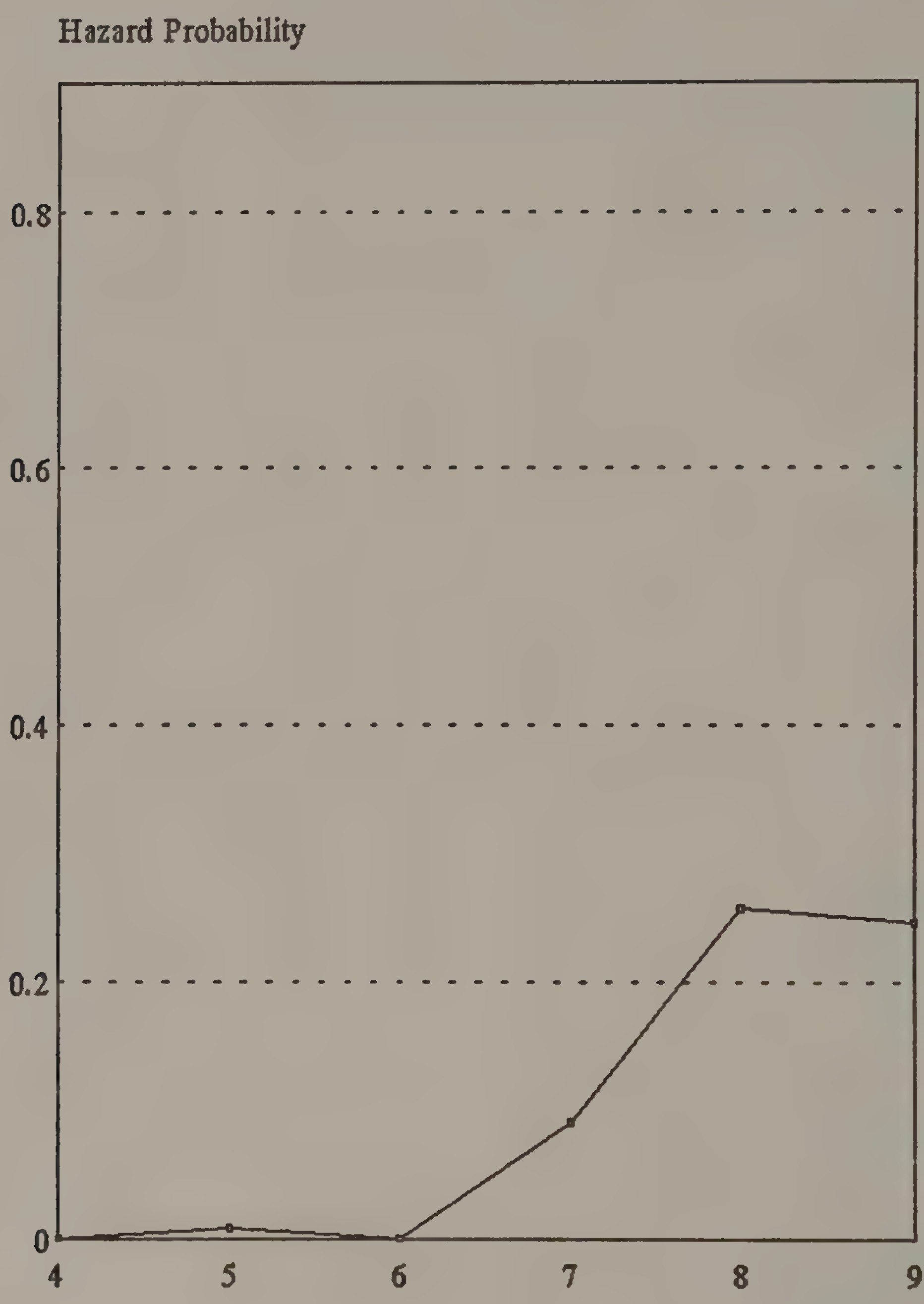
	Full Sample	Households with Daughters In-School	Households with Daughters Out-of-School	T statistics or χ^2	
Family structure					
Percentage of mothers married	54	52	56	0.13	χ^2
Percentage of fathers full-time residents of households with husbands	55	50	59	0.52	χ^2
Percentage of fathers full-time residents of all households	30	26	33	0.63	χ^2
Mother's age (mean)	45.0	43.1	46.8	2.25*	t
Mean age range of children at home	14.1	12.8	15.3	2.08*	t
Mean age children age 6-18 years	11.0	11.4	10.6	-1.19	t
Family economy					
Labor demand					
Percent employed in wage sector, mother	30	36	25	1.90	χ^2
Percent employed in wage sector, senior male if there is one	34	39	28	1.80	χ^2
No. daily household work tasks, mother	5.8	6.0	5.6	-0.52	t
No. daily household work tasks, mother and daughter together	4.2	4.0	4.5	1.06	t
Gender gap: daughter minus son tasks	0.1	-0.2	0.4	2.24*	t
Market production (% yes by item)					
traditional beer	36	31	41	1.28	χ^2
cattle	31	31	31	0.00	χ^2
goats	33	31	34	0.15	χ^2
chickens	22	18	26	1.19	χ^2
grains	16	21	11	2.15	χ^2
(Volume, median number) of sacks					
sorghum	3	5	1		
maize	0	0	0		

TABLE 1: Characteristics of Botswana Households, Mothers, and Daughters (Continued)

	Full Sample	Households with Daughters In-School	Households with Daughters Out-of-School	T statistics or χ^2	
Family economy (continued)					
Livestock assets (Median number owned)					
Cattle	1	0	4		
Chickens	5	5	4		
Goats	9	10	6		
Consumer goods (% yes by item)					
Radio	75	69	80	2.12	χ^2
Shoes	99	98	100	1.03	χ^2
Bicycle	34	31	36	0.33	χ^2
Tools/productive goods (% yes by item)					
Plough	60	60	59	0.01	χ^2
Donkey cart	32	30	34	0.34	χ^2
Donkeys	37	36	38	0.04	χ^2
Motor vehicle	18	30	7	10.87**	χ^2
Household quality and investments (% yes)					
Square house in compound	78	89	66	8.65**	χ^2
Glass windows	83	90	75	4.85*	χ^2
Water standpipe	24	40	8	16.79***	χ^2
Primitive latrine	10	13	7	1.48	χ^2
High quality thatch roof	68	70	66	0.34	χ^2
Schooling, literacy, and social beliefs					
Mother's years of schooling	4.4	5.4	3.5	-2.76**	t
Target daughter's years of schooling	8.5	9.0	8.1	-7.85***	t
Senior male's years of schooling	5.1	5.1	5.2	0.09	t
Mother's literacy (10 items possible)	7.2	8.1	6.3	-3.35***	t
Daughter's literacy (7 items possible)	3.4	4.5	2.3	-8.63***	t
Perceived quality of local school	22.0	23.2	20.5	-3.20**	t
Percentage of mothers using birth control method	42	55	30	8.06**	χ^2
N	122	61	61		

T statistics or χ^2 * $p < .05$ ** $p < .01$ * $p < .001$

FIGURE 1: Baseline Hazard Probability of Girls Leaving School in Botswana



(N = 122)

TABLE 2: Influence of Family Economic Factors on the Probability of Daughters Leaving School[a]

Predictors	Model A	Model B	Model C
Grades 1-4	—	—	—
Grade 5	-4.80**	-5.17***	-6.67***
	(1.00)	(1.44)	(1.76)
Grade 6	—	—	—
Grade 7	-2.30**	-2.64**	-3.92**
	(0.32)	(1.07)	(1.46)
Grade 8	-1.06**	-1.24	-2.21
	(0.22)	(1.04)	(1.42)
Grade 9	-1.12**	-1.21	-2.13
	(0.26)	(1.05)	(1.42)
Household quality and investment		-0.41***	-0.16
		(0.10)	(0.13)
Household production		0.07	0.07
		(0.06)	(0.07)
Consumer goods		0.13	0.12
		(0.11)	(0.11)
South villages		0.74*	3.26***
		(0.35)	(0.97)
Mother's employment			-0.01
			(0.36)
Senior male's employment			0.14
			(0.37)
Gender gap in labor tasks demanded of daughter minus son			0.01
			(0.14)
Household quality * South			-0.59**
			(0.21)
Gender gap in tasks * South			0.39*
			(0.20)
2 log likelihood	300.08	282.41	264.84
Change in log likelihood		17.67***	17.57**
Change in df		4	5

[a] Hazard functions, parameter estimates, and standard errors reported.

χ^2 * $p < .05$ ** $p < .01$ *** $p < .001$

Family-Economy Effects

The basic family-economy model is moderately efficacious in explaining variation in whether and when girls leave school. Just one of the three measures of household wealth proved to be related to daughters' school attainment: the index of household quality and investment that is added to the baseline hazard-rate function in model B (Table 2). The decrement in the chi-square value is highly significant. The coefficient's negative sign indicates that the probability of dropping out is *lower* among families with higher scores on the household-quality index. The composite measures of household production and ownership of consumer goods show no effect.

Above we saw wide variation in the number of household tasks required of children, and unequal labor demands placed on daughters versus sons. But neither household-labor nor employment-status predictors held significant main effects on daughters' attainment. The level of gender inequality found within southern villages does significantly *raise* the probability of girls' dropping out. This demonstrates the utility of broadening the standard family-economy model to look at gender-specific labor demands. A direct trade-off between time spent in school and working may not exist for most Botswana youths. In general, many children and youths appear to have a good deal of discretionary time, not found in peasant societies that are more dependent upon subsistence agriculture.[10]

Next we turn to the wider set of social factors specified in our *family-institution* model that may independently influence daughters' school attainment or which my interact with family-economy factors. Social and institutional processes are split into two areas: family social structure and mothers' own schooling, literacy, social beliefs, and practices.

Family Social Structure

Two elements of family structure are significantly related to daughters' attainment. First, girls with older siblings generally stay in school longer, as evidenced by the effects from average age of children in the household (Table 3, model B). Daughters with siblings of average age, 11 years, were 3.5 times more likely to still be in school, compared to daughters with siblings of mean age, 5 years. Our qualitative work in one village suggests that this effect is due to more intense child-care responsibilities of daughters who have fewer older siblings (Fuller et al. 1992).

Second, mothers and daughters residing in the two northern villages had significantly *higher* educational attainment — even though these villages had a higher prevalence of absentee fathers. Only 19% of mothers in the North reported that the daughter's father resided in the household, versus 40% among mothers in the South. The family economies looked quite similar, with only sorghum production (for market) higher in the villages in the South than in the North. However, northern mothers had 5.2 years of schooling versus 3.7 years among southern mothers. Daughters had 8.7 years of schooling in the North, compared to 8.4 years among daughters in the South. It may be that labor out-migration is more frequent from the North where proximity to urban labor markets is more distant. This does not necessarily disadvantage the family

economically; women tend to have *higher* school attainment in father-absent households. From at least one other developing country, evidence suggests that mothers support their daughters' nutrition and schooling more equitably when the father is no longer in the household (Schultz 1990). In Botswana, 34.9% of all farming households are headed by women (Botswana Central Statistics Office 1989). Other facets of family structure held no discernible influence on daughters' attainment: family size, number of resident adults, and males' education level.

Literacy Practices, School, and Fertility Beliefs

In contrast to the inconsistent effects of the family economy, several social commitments and practices exercised by the mother were directly related to the daughter's attainment. In Table 3 we report on effects from mother's literacy score, perceived level of school quality, and use of a birth control device. All three social commitments reduce the log likelihood of dropping out when entered sequentially. Girls with mothers who do not use birth control are twice as likely to leave school. Girls whose mothers scored in the highest literacy decile, on average, had entered grade 9, compared to median school attainment of grade 7 for girls whose mothers scored in the lowest decile.[11]

The mother's perceived quality of the local school and the general utility of modern schooling also were associated with a lower probability of dropping out (models C and D). Linear and nonlinear (squared) relationships must be interpreted together (in model D). When we plot the overall hazard function, the effect of mother's literacy on *reducing* the probability of the daughter's dropping out is significant (see Figure 2); the positive effect of mother's school attainment alone is marginal. A similar plot (not shown here) of linear and nonlinear effects of the school's utility shows a significant reduction in the risk of dropping out. The mother's preference for a more obedient (less liberal) daughter *raises* the probability that the daughter will leave school (index detailed in Appendix). Overall, note the substantial reduction in the -2 log likelihood between model A and model D (Table 3) when all social factors are in the equation; this reduction is much less for the family-economy model (Table 2, model C).

Do mothers' social commitments interact with family-economy factors? We tested for effects from three possible interaction terms between the household-investment indicator (wealth) and social-commitment variables. The interaction between investment and mothers' perception of school quality did significantly reduce the chi-square ($p < .01$).

Fully Interactive Grade-Level Models

The next analytic step (1) directly tests the influence of social factors, after controlling for the influence of family-economy factors, and (2) assesses whether economic and social variables interact with school grade levels. The latter procedure provides an assessment of whether the school's own institutional structure interacts with family-level economic and social forces. For example, the end of primary school (grade 7) may be a point at which some parents (with specified economic or social attributes) allow their daughter to leave school.

TABLE 3: Influence of Family Social Variables on the Probability of Daughters Leaving School[a]

Predictors	Model A	Model B	Model C	Model D
Grades 1-4	—	—	—	—
Grade 5	-4.80** (1.00)	-2.61* (1.09)	—	—
Grade 6	—	—	—	—
Grade 7	-2.30** (0.32)	0.05 (0.56)	2.16* (0.99)	-17.91 (10.25)
Grade 8	-1.06** (0.22)	1.35* (0.56)	3.59** (1.04)	-16.46 (10.25)
Grade 9	-1.12** (0.26)	1.49* (0.61)	3.40** (1.07)	-16.34 (10.24)
Average age of children		-0.16** (0.05)	-0.17** (0.07)	-0.25** (0.08)
Mother's literacy level		-0.12* (0.05)	-0.12* (0.05)	-0.14 (0.08)
Use of birth control		-0.75* (0.33)	-0.67 (0.36)	-0.91* (0.41)
Perceived quality of local school			-0.10** (0.04)	-0.13** (0.04)
Mother's school attainment level				0.35* (0.18)
Mother's attainment, nonlinear component				-0.03* (0.02)
Mother's preference for obedient daughter				1.44* (0.64)
Daughter obedience, nonlinear component				-0.56* (0.25)
Perceived usefulness of school				3.93* (1.83)
Usefulness of school, nonlinear component				-0.18* (0.08)
-2 log likelihood	300.08	272.05	215.45	182.50
Change in log likelihood		28.03***	56.60***	32.95***
Change in df		3	1	6

[a] Hazard functions, parameter estimates, and standard errors reported.

χ^2 * p < .05 ** p < .01 *** p < .001

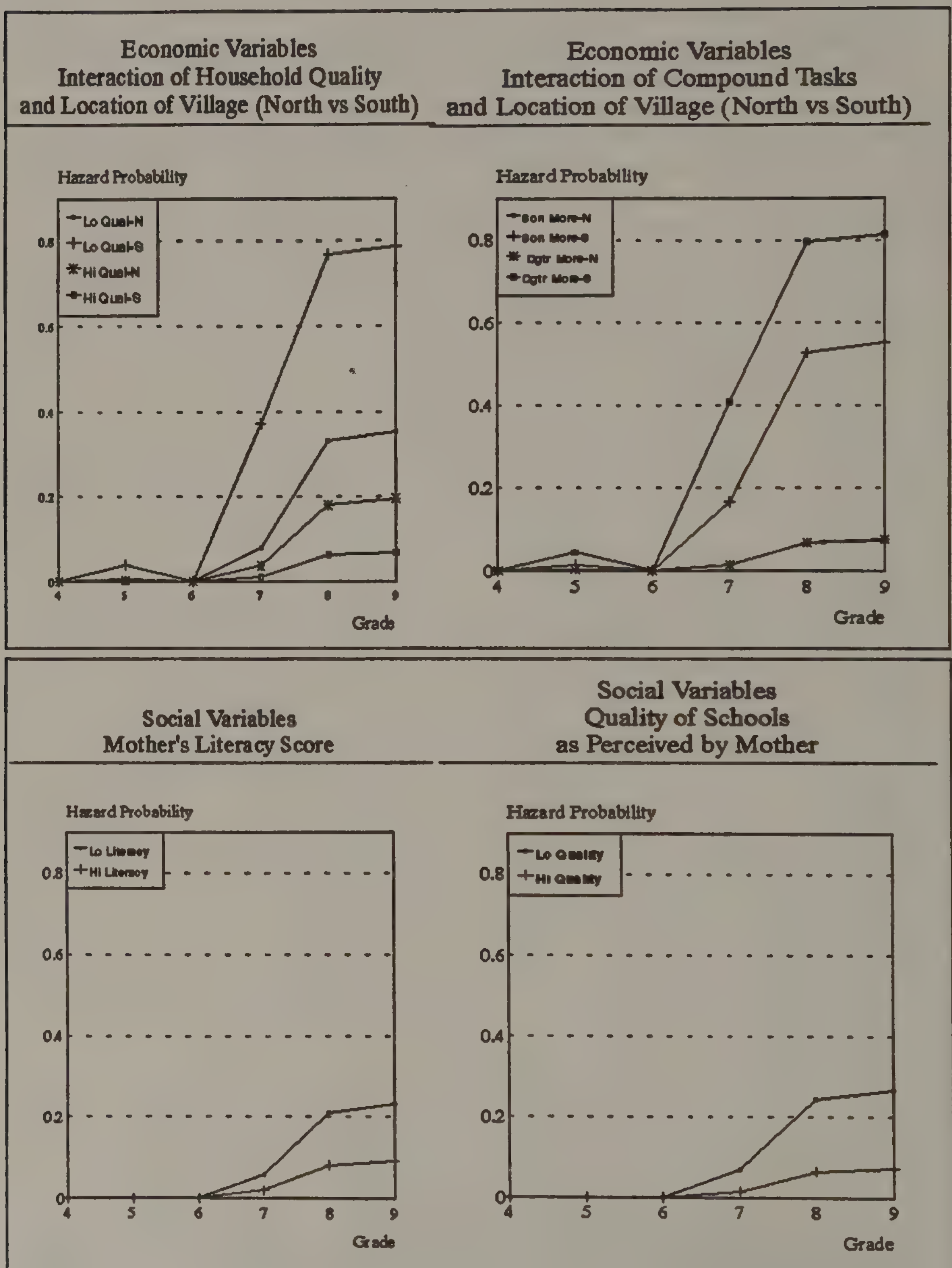
Where earlier predictors were found to have a significant interaction effect with any *one* grade level on the probability of leaving school, interaction terms between these predictors and *all* grade levels (grade 5-9) were entered.[12] To summarize these findings (the detailed statistical table is available from the authors), first for interactions between specific school grade levels and family-economy factors: the probability of dropping out is higher at the end of grade 7 in households with a greater gender gap in labor tasks. Daughters in the South also have a higher drop-out risk during grades 8 and 9, perhaps when labor demands exerted by the wage economy begin to be felt. Interaction terms for grade levels and maternal social factors are then entered, controlling on the family-economy variables. In addition to already mentioned main effects, the perceived quality of the school holds an effect, particularly during grades 7 and 9. Household quality, mother's literacy, and use of birth control all remain important factors in lowering the drop-out risk, *independent of the daughter's grade level* (i.e., no interaction effects).

In sum, elements of the family-institution model explain additional proportions of variance in daughters' drop-out risk: age structure of the household and implied child-care demands placed on daughters, and particularly, maternal literacy, fertility practices, and child-rearing views (liberal or traditional). The mother's perceptions of schooling may interact with institutional realities in terms of how she perceives the quality of the local school and the utility of formal education in general. And discrete institutional effects are seen in the specific (terminal) grade levels at which the drop-out risk is greatest.

The *magnitude* of selected interaction effects is illustrated in Figure 2. For the family-economy model (Table 2, model C) two interactions have been plotted. First is the interaction of the village's southern location and household quality. The estimated odds of leaving school in any given year are always higher if the assessed quality of the compound is low; but these effects, holding other variables constant, differ depending on whether the compound is in a southern or northern village. In low-quality households in the South, the drop-out risk is over twice as great as in low-quality households in the North. Similar interaction effects are seen in the second economic plot, illustrating the influence of village location and difference in the number of household tasks for which the daughter and son are responsible. In the South, daughters who perform more household tasks than sons are one and a half times as likely to drop out of school than daughters who do fewer household tasks than sons; whereas, in the North, there is little apparent drop-out risk.

For the full family-institution model (Table 3, model D) six predictors are highly related to drop-out risk. Thus the effect of any particular predictor appears to be smaller. Two of the variables have been chosen to illustrate the magnitude of these main effects. The first (Figure 2, bottom left) depicts the effect of mother's literacy score on drop-out risk. In households where the mother has attained a high score on the literacy test, daughters are two and a half times less likely to leave school than daughters whose mothers score low on the test, holding other family-institution factors constant.

FIGURE 2: Hazard Probabilities of Girls Leaving School in Botswana



(N = 122)

Conclusions

At first glance, our findings might be interpreted as demonstrating the narrowness of the family-economy model, at least when studying female attainment in this African setting. Indeed, *gender-specific* facets of the household economy and family institution — especially maternal commitments related to fertility, child rearing, and schooling — are more influential in explaining daughters' propensity to leave school than more generic facets of family-economy. Our findings, however, do confirm two particular elements of the basic family-economy framework. First, local economic conditions are setting levels of discretionary time available to young females. Just as Walters and O'Connell (1988) found that nineteenth-century rural children in North America had more discretionary time to attend school than urban youths who responded to industrial labor demands, we found sizable amounts of discretionary time. This minimizes the direct trade-off between daughters' work demands and time available for school. Plentiful discretionary time for *mothers* (due to cash remittances from family members) may also condition the variability and attainment effects of maternal social practices.[13]

Second, we find that culturally specific indicators of class position and household resources are related to daughters' attainment levels, especially the family's propensity to save and invest in the quality of their compound. In contrast, neither spending on nondurable consumer goods nor possession of productive assets was related to daughters' attainment. How families accumulate and spend their income may indicate an underlying commitment to longer-term investments. Our findings also lend support to Schultz's (1990) finding that mothers, when they control the allocation of family resources, tend to support their daughters' schooling more equitably relative to that of their sons. We observe this effect in the two most southern villages, where access to jobs in South Africa is closer for husbands. Comparative-historical scholars, data permitting, might pursue this issue of how mothers mediate labor demands and influence the allocation of household resources.

Finally, our findings demonstrate the importance of broadening current conceptions of how the family operates within pluralistic institutional environments, including modernizing contexts that manifest alternative social beliefs and economic options. Maternal commitment to literacy practices at home and their perceptions of higher school quality were strongly related to daughters' achievement, after controlling on family-economy effects. More general beliefs, indicated by endorsement of birth control, also are related to daughters' attainment. This fits with work from Kenya, Mexico, and Nepal showing that more educated mothers individualize their child-rearing practices and explicitly attempt to "develop" their child (LeVine 1987). Precisely how comparative-historical research can get at variation in parents' social commitments, and within what institutional settings these factors interact with family-economy processes, remain intriguing analytic questions.[14] Our analysis also showed interaction effects from terminal grade levels (at the end of primary and junior-secondary school), again evidence of how external institutions blend with household-level forces to influence girls' school attainment.

Proponents of the family-economy framework are pushing us to look into the basic location of action from which parental demand for schooling emanates and variable levels of child attainment result. It moves beyond the assumption that all families respond to the macrostructure in uniform ways; it recognizes that "choices" available to the household are constrained by opportunities set by institutions and markets. By integrating the strengths of this theoretical development with emerging work on parents' social commitments within contemporary settings, we can begin to build a more complete picture of the family organization as situated within a variegated institutional environment. Families, especially during periods of commercialization or symbolic modernization, look outside and confront competing economic demands, social commitments, and parenting practices. In turn, how mothers extract these economic resources and normative convictions from these diverse environs influences how far their daughters go in school.

Notes

1. These conceptual advances are important, particularly in addressing Coleman's (1986) criticism that we often infer local forms of, and motives for, agency from the institutional processes that are deduced from macro theory — but which are rarely observed directly. Avoiding the adoption of an idealized conception of voluntaristic, intention-filled agency, family-economy theory is beginning to demonstrate how the family acts with bounded rationality, given broader institutional constraints and alternative scripts. To illuminate the novel character of the family-economy model, one can contrast it with the status-attainment tradition, focusing on individual attributes of youths (for review, Campbell 1983), and with macrostructural explanations, focusing on historical change in state action, school supply, and institutions (for review, Fuller & Rubinson 1992).

2. Early work on modernization, for example, examined variation among parents in their beliefs and commitment to broad Western values: individualism versus conformity, technical innovation versus traditional farming or work practices (e.g., Inkeles & Smith 1974). Early school-attainment research in the U.S. also examined the influence of parents' expectations and aspirations regarding their child's education (Campbell 1983). But in recent competitive tests between generalized expectations and specific parental strategies (supervision of homework, contacts with school, for instance), the latter proximate causes have been found to be more influential (Astone & McLanahan 1991; Dornbusch et al. 1985). Parental beliefs toward the school's legitimacy and importance can vary somewhat independently of such parenting practices, and may vary within social-class groups (for Japan and the U.S., see Holloway et al. 1990). Variation in cultural or religious commitments to schooling *within* particular social classes also may help to explain interfamily variability in socialization beliefs and parents' push on their children for higher attainment (in the U.S., Alwin 1984; in Africa, Hyde 1993).

3. Initial evidence shows that culturally relevant measures of family social status more consistently predict school achievement, compared to conventional indicators of SES borrowed from Western conceptions of class (for review, Fuller & Clarke 1994). And note that historical research on rural families in the U.S. at times imputes certain social meanings from particular measures of "wealth." For example, higher levels of school attainment among families with more mechanized farming practices are associated with a Calvinist commitment to long-term investment and technological improvement (Guest & Tolnay 1985). But could we make the same inference from economic indicators linked to consumption of nondurable goods or higher levels of farm productivity? Multiple measures of wealth, informed by the culturally specific meaning of each, would reveal more about the consumption preferences of families.

4. Similarly, work in the U.S. shows that adolescent daughters are provided more social opportunities when the mother's class position and influence are more equal to the father's (Hagan, Simpson & Gillis 1987).

APPENDIX: CONSTRUCTED INDICES

Most measures of economic and social variables are discrete and self-explanatory. Construction of several composite indices was done as described below. The interview instrument, available from the authors, was field-tested in two of the four villages.

Economic Indices

A checklist of 27 frequently performed work tasks was utilized, including employment in the wage sector, work on cattle lands or farming outside the household, and domestic tasks inside the family compound. For each task, the mother indicated who normally performed the task: the mother, daughter, husband or senior male, or son. More than one family member could be indicated.

The index of *household production* contained the dichotomous values for crops or livestock most commonly produced for, and taken to, market (see Table 1). Elements of the *consumer goods* and *household quality* indices also are listed in Table 1. Commonly understood criteria were used to distinguish high- versus low-quality thatched roofs and latrines.

To identify culturally relevant work tasks and economic indicators, prior to the field tests we consulted earlier ethnographic research and recent household surveys in Botswana (Alverson 1978; Kossoudji & Mueller 1983; Cownie & Blake 1989).

Social Indices

Mothers' perceptions of *school quality* were recorded by asking nine questions pertaining to the perceived quality of teachers, the school principal's responsiveness to the community, and the quality of the school grounds. Each item used a five-point scale, from "strongly agree" to "strongly disagree."

Four specific questions were utilized to construct the *usefulness of school* index. These items ranged from "Schooling will help my daughter find a wage-earning job" to "Schooling for my daughter will mean she will pass more money to her in-laws and not to me when she gets married" (a frequently mentioned reason for not letting daughters go "too far" in school).

We asked mothers about the behaviors that they like to see in their daughters. Six dichotomous choices were given to the respondent: "Which behavior do you prefer — obeys me *or* learns to speak-up if she disagrees with me?" We constructed the *preference for an obedient daughter* index by simply summing the dichotomous values.

---

5. Each village selected has a junior secondary school within walking distance. Our research team (two interviewers) started at the school, randomly selecting the 15 in-school girls. Each female student was accompanied home. The mother was interviewed; brief literacy tests were given to the mother (orally) and to the daughter (pencil and paper). A bottle was then spun in front of this mother's compound, and we walked in the direction indicated until a mother was found with an eligible out-of-school daughter. This procedure was not intended to yield a random sample: nationwide 65% of all girls in this age cohort are still enrolled in school. Our purposeful sampling technique aimed, first, to provide contrasting subsamples of in-school and out-of-school girls who lived within the same sections of the four villages (Mochudi and Moshupa in the south, close to the capital city and the South African border; Bobonong and

Mmadinare in the north, farther from urban labor markets). The villages were selected primarily for their variable proximity to wage labor markets. Second, we wanted to ensure that distance to school was not the principal explanatory factor. Each village had a junior secondary school close to the village center. We controlled on distance-to-school by first locating a daughter currently attending school, then circling around that household until we located a family with a similarly aged, but out-of-school daughter. All four villages are populated by members of the dominant Tswana ethnic group.

6. Years of school attainment was highly correlated with actual literacy scores ($r = .49, p < .001$). In regressing daughters' literacy scores (based on our 12-item reading comprehension exam) on years of school attained, mother's literacy, and seven additional social and economic household variables, daughter's attainment remained highly significant. The beta coefficient indicated that one year of additional schooling was associated with one additional item correct. The coefficient was five time the standard error. These findings are detailed in an earlier paper (Fuller et al. 1992).

7. Household economists always use individual-level constructs to model school attainment, since they assume that original agency sprouts from this level. Reviewed above, comparative-historical scholars tend to focus on aggregates, for instance, labor demand within a local area or aggregate family income. Studies of family demand or school attainment in contemporary settings often use such contextual measures as well as variables pertaining to individual family members (Horan & Hargis 1991; Rosenzweig & Schultz 1982).

8. Contemporary evidence demonstrates covariation between mother's schooling, the frequency of mother-child interaction, explicit socialization of the individual child (versus more implicit and collective upbringing of the child in the village), and lower fertility rates (for review, LeVine 1987).

9. Studies of enrollment demand in the comparative-historical tradition rely on decennial census or annual survey data. We already know that the influence of contextual economic or political factors on school demand depends upon the time lag used in time-series analyses (Fuller et al. 1992; Walters & Rubinson 1983). As this line of work moves closer to how family characteristics interact with local contexts, time-dependencies linked to whether and when children leave school should receive more specific attention.

10. Tswana settlements traditionally gained cash from cattle for production, which limited labor demands on women and children who remained in the village household. Throughout the nineteenth century, labor migration of males to South African mines and other wage employment helped to raise cash income while making female-headed households very common. In 1971, for instance, there were just 63 young men residing in households for every 100 women in the 20-39 age cohort (Brown 1983). One time-use study in Botswana found that up to 40% of rural adults' time is spent in leisure activities (Kossoudji & Mueller 1983).

11. In 1988, 24% of all young females who had completed some period of secondary schooling, age 15-19, had experienced one pregnancy. This rate of child-bearing was 33% for young females who had not completed primary school (Cownie & Blake 1989).

12. These models become complex. But we have an abundance of degrees-of-freedom, given that the person-period data set, at a minimum, contains over 450 records, one for each person-year of data.

13. Household economists *assume* that parents can vary the amount of time they devote to raising children (Becker, Murphy & Tamura 1990). Note that the Botswana case indicates that discretionary time itself is conditioned by local economic and institutional factors.

14. Historical studies of school demand are taking small steps, for example, beginning to look at the influence of Protestant or Catholic church membership as one indicator of a modern commitment to liberal forms of child development and status attainment (Boli et al. 1985).

References

Alverson, Hoyt. 1978. *Mind in the Heart of Darkness.* Yale University Press.

Alwin, Duane. 1984. "Trends in Parental Socialization Values: Detroit, 1958-1983." *American Sociological Review* 49:359-82.

Angel, Ronald, and Marta Tienda. 1982. "Determinants of Extended Household Structure: Cultural Pattern or Economic Need?" *American Journal of Sociology* 87:1360-83.

Astone, Nan Marie, and Sara McLanahan. 1991. "Family Structure, Parental Practices and High School Completion." *American Sociological Review* 56:309-20.

Baker, David, and David Stevenson. 1986. "Mothers' Strategies for Children's School Achievement." *Sociology of Education* 59:156-66.

Becker, Gary. 1976. *A Treatise on the Family.* Harvard University Press.

Becker, Gary, K.M. Murphy, and R. Tamura. 1990. "Human Capital, Fertility and Economic Growth." *Journal of Political Economy* 98:512-37.

Boli, John, Franciso Ramirez, and John W. Meyer. 1985. "Explaining the Origins and Expansion of Mass Education." *Comparative Education Review* 29:145-70.

Botswana Central Statistics Office. 1989. *1988 Botswana Agricultural Statistics.* Gaborone.

Brown, Barbara. 1983. "The Impact of Male Labour Migration on Women in Botswana." *Journal of the Royal African Society* 82:367-85.

Campbell, Richard T. 1983. "Status Attainment Research: End of the Beginning or Beginning of the End?" *Sociology of Education* 56:47-62

Carnoy, Martin, and Henry Levin. 1985. *Schooling and Work in the Democratic State.* Stanford University Press.

Colclough, Christopher, and Stephen McCarthy. 1980. *The Political Economy of Botswana.* Oxford University Press.

Coleman, James S. 1986. "Social Theory, Social Research, and a Theory of Action." *American Journal of Sociology* 91:1309-35.

Cownie, David, and Elizabeth Blake. 1989. *Children, Women and Development in Botswana.* Gaborone: UNICEF.

Dixon-Mueller, Eva. 1985. *Women's Work in Third World Agriculture.* Geneva: International Labor Organization.

Dornbusch, Sanford, J. Merrill Carlsmith, Steven Bushwall, Philip Ritter, Herbert Leiderman, Albert Hastorf, and Ruth Gross. 1985. "Single Parents, Extended Households, and the Control of Adolescents." *Child Development* 56:326-41.

Duncan, Wendy. 1989. *Engendering School Learning.* Stockholm: Institute of International Education.

Fuller, Bruce, and Prema Clarke. 1994. "Raising School Effects while Ignoring Culture?" *Review of Educational Research* 64:119-57.

Fuller, Bruce, Jerald Hage, Maurice Garnier, and Max Sawicky. 1992. "Nation Building and School Expansion under the Fragile French State." *Social Forces* 70:923-36.

Fuller, Bruce, Margaret Keiley, Judith D. Singer, and David Cownie. 1992. "The Family's Influence on Girls' School Attainment: Evidence from Botswana." Paper presented at American Educational Research Association, San Francisco.

Glewwe, Paul. 1991. "Student Achievement and Schooling Choice in Low Income Countries: Evidence from Ghana." Washington: World Bank (draft manuscript).

Goodnow, Jacqueline, and W. Andrew Collins. 1990. *Development According to Parents: The Nature, Sources, and Consequences of Parents' Ideas.* Erlbaum.

Guest, Avery, and Stewart Tolnay. 1985. "Agricultural Organization and Educational Consumption in the U.S. in 1900." *Sociology of Education* 58:201-12.

Hagan, John, John Simpson, and A.R. Gillis. 1987. "Class in the Household: A Power-Control Theory of Gender and Delinquency." *American Journal of Sociology* 83:788-816.

Hareven, Tamara. 1982. *Family Time and Industrial Time: The Relationship between the Family and Work in a New England Industrial Community*. Cambridge University Press.

Hauser, Robert, and Thomas Daymont. 1977. "Schooling, Ability, and Earnings: Findings 8 to 14 Years after High School Graduation." *Sociology of Education* 50:182-206.

Hernes, Gudmund. 1976. "Structural Change in Social Processes." *American Journal of Sociology* 82:513-47.

Hess, Robert, and Susan Holloway. 1984. "Family and School as Educational Institutions." Pp. 179-222 in *Review of Child Development Research*. Vol. 7, edited by Ross Parke. University of Chicago Press.

Hill, M.Anne, and Elizabeth King. 1993. "Women's Education in Developing Countries: An Overview." Pp. 1-50 in *Women's Education in Developing Countries: Barriers, Benefits, and Policies*, edited by E. King and A. Hill. Johns Hopkins University Press.

Holloway, Susan, Bruce Fuller, Robert Hess, Hiroshi Azuma, Keiko Kashiwagi, and Kathleen Gorman. 1990. "The Family's Influence on Achievement in Japan and the United States." *Comparative Education Review* 34:196-207.

Horan, Patrick, and Peggy Hargis. 1991. "Children's Work and Schooling in the Nineteenth-Century Family Economy." *American Sociological Review* 56:583-96.

Hyde, Karin. 1993. "Sub-Saharan Africa." Pp. 100-35 in *Women's Education in Developing Countries: Barriers, Benefits, and Policies*, edited by Elizabeth King and M. Anne Hill. Johns Hopkins University Press.

Inkeles, Alex, and David H. Smith. 1974. *Becoming Modern*. Heinemann.

Kahn, Michael. 1988. "Indicators of Equity in Junior Secondary Education in Botswana." Pp. 119-36 in *Botswana: Education, Culture, and Politics*, edited by Ansu Datta and Kenneth King. University of Edinburgh.

Knight, John, and Richard Sabot. 1990. *Education, Productivity, and Inequality: The East African Natural Experiment*. Oxford University Press.

Kohn, Melvin, Atsushi Naoi, Carrie Schoenbach, Carmi Schooler, and Kazimierz Slomczynski. 1990. "Position in the Class Structure and Psychological Functioning in the United States, Japan, and Poland." *American Journal of Sociology* 95:964-1008.

Kossoudji, Sherrie, and Eva Mueller. 1983. "The Economic and Social Status of Female-headed Households in Rural Botswana." *Economic Development and Cultural Change* 31:825-38.

LeVine, Robert. 1987. "Women's Schooling, Patterns of Fertility, and Child Survival." *Educational Researcher* 16:21-27.

Ogbu, John. 1978. *Minority Education and Caste: The American System in Cross-cultural Perspective*. Academic Press.

Robinson, Robert. 1993. "Economic Necessity and the Life Cycle in the Family Economy of Nineteenth-Century Indianapolis." *American Journal of Sociology* 99:49-74.

Rosenzweig, M.R., and T. Paul Schultz. 1982. "Market Opportunities, Genetic Endowments, and Intrafamily Distribution: Child Survival in Rural India." *American Economic Review* 72:803-15.

Rubinson, Richard. 1986. "Class Formation, Political Organization, and Institutional Structure: The Case of Schooling in the United States." *American Journal of Sociology* 92:519-48.

Schultz, T. Paul. 1990. "Testing the Neoclassical Model of Family Labor Supply and Fertility." *Journal of Human Resources* 25:599-635.

Schultz, Theodore. 1961. "Investment in Human Capital." *American Economic Review* 51:1-16

Sewell, William, and Robert Hauser. 1975. *Education, Occupation, and Earnings*. Academic Press.

Singer, Judith D., and John B. Willett. 1993. "It's About Time: Using Discrete Time Survival Analysis to Study Duration and the Timing of Events." *Journal of Educational Statistics*. 18:155-95.

Tilly, Louise, and Joan Scott. 1978. *Women, Work and Family*. Holt Rinehart.

Tuma, Nancy, and Michael Hannan. 1984. *Social Dynamics: Models and Methods*. Academic Press.

Walters, Pamela Barnhouse. 1992. "Who Should Be Schooled? The Politics of Class, Race, and Ethnicity." Pp. 173-87 in *The Political Construction of Education*, edited by Bruce Fuller and Richard Rubinson. Praeger.

Walters, Pamela B., and David R. James. 1992. "Schooling for Some: Child Labor and School Enrollment of Black and White Children in the Early Twentieth Century South." *American Sociological Review* 57:635-50.

Walters, Pamela Barnhouse, and Philip O'Connell. 1988. "The Family Economy, Work, and Educational Participation in the United States, 1890-1940." *American Journal of Sociology* 93:1116-52.

Walters, Pamela Barnhouse, and Richard Rubinson. 1983. "Educational Expansion and Economic Output in the United States, 1890-1969." *American Sociological Review* 48:480-93.

World Bank. 1991. *World Development Report*. Washington.

Religion, Health, and Nonphysical Senses of Self*

ELLEN L. IDLER, *Rutgers University*

Abstract

The relationship between religion and physical health is a complicated one. In the Durkheimian tradition, and in longitudinal epidemiological studies, religious involvement is shown to have a protective effect on health. Cross-sectionally, however, and even in short followup periods, we sometimes see higher levels of religiousness associated with poorer, not better, health, as people in the midst of crises often turn to religion for comfort and social support. A third way of thinking about the relationship involves self-ratings of health, which appear to represent broad conceptions of self in which physical health and abilities may be deemphasized and nonphysical characteristics, including religious or spiritual self-identities, may be relied upon. Quantitative and qualitative data from a cross-sectional sample of disabled clients of an urban rehabilitation clinic support both of the latter perspectives.

Social scientists in the past several decades, who have been very much interested in health and social life, have generally shown very little interest in the subject of religion and health. For one thing, data sets in which both these complex phenomena are adequately measured are rare; even rarer are followup studies in which causal direction can be accurately assessed. The primary source of funding for health research in the social sciences has been the federal government and, although no explicit restrictions are placed on survey research on religious practices, a de facto restriction has arisen from the principled separation of church and state. The comparative lack of research on religion may also reflect some absence of interest on the part of health researchers. Another reason may be that the relationship between religion and health is so

** An earlier version of this article was presented at the meetings of the Southern Sociological Society, Raleigh, North Carolina, April 1994. The research reported here was supported by the National Institute on Aging FIRST award (AG-07480-02). I would like to thank Shelley Myer and Sharon Cook for their assistance with the data collection and analysis; Linda George and Allan Horwitz and two anonymous reviewers for their comments on an earlier draft of the article; and the physicians, nurse, and most especially the clients of the disability clinic for their candid and thoughtful answers to my questions.* *Address correspondence to Ellen L. Idler, Institute for Health, Health Care Policy, and Aging Research, 30 College Avenue, New Brunswick, New Jersey, 08903.*

 Social Forces, December 1995, 74(2):683-704

complicated. Alternative causal orders are equally plausible, and even the sign of the association is uncertain. *The Oxford English Dictionary* says "complicated" means that which has been "folded, wrapped, or twisted together, combined intimately, and mixed up with in an involved way." This describes the relationship of religion and health very well. Moreover, neither religion nor health is itself a single strand. Religious involvement can mean a mix of practices, beliefs, and identities; health is an even more global concept combining mental, physical, and even social well-being. The purpose of this article is not really to untwist the two, since they appear to occur as naturally interwoven, but simply to look at the knot from several vantage points.

View 1: Religion's Protective Effect on Health

One way to think about religion and health is the way Durkheim ([1897] 1951; [1915] 1965) did. Religion has a beneficial effect on human social life and individual well-being because it regulates behavior and integrates individuals in caring social circles. It provides stability and support, structure and intimacy. Ritual religious practices especially provide these benefits, and religious faiths with high levels of ritual in their worship and private devotional practices tend to benefit more from lower suicide rates (Durkheim [1897] 1951) and better health among the elderly (Idler & Kasl 1992). Even more directly, religious belief systems may include prohibitions against certain behaviors such as smoking, excessive alcohol use, or sexual experimentation, lowering the incidence among religious group members of diseases for which these behaviors are risk factors (Levin & Schiller 1987). Less directly, religious beliefs may provide an overarching cognitive framework for interpreting events, "plausibility structures" that are readily available to individuals to provide coherent meaning for both the small events of the life course and the large events of history (Berger 1967). These mechanisms have sometimes been called the health behavior, social cohesiveness, and sense of coherence hypotheses (Idler 1987; see also Idler 1993a). Because of its multidimensionality, religious involvement could have multiple causal pathways for its effect on physical health.

The causal order in this view is that religious involvement precedes health status; the direction of the association is positive. As in any research on the effect of risk factors on health, the most conclusive research comes from epidemiological studies of large populations with extensive initial health assessments and longitudinal followup of subjects, just the kind of data that do not usually include measures of religious involvement, let alone richly multidimensional ones.

View 2: Religion in Illness and Suffering

Paradoxically, the mirror image of the relationship between religion and health also has a compelling logic. People often turn to religion in times of trouble, including, and especially, during crises of serious illness. Religious groups can offer both spiritual and practical help to the sick: prayers, visitors, and hot meals. Perhaps even more important, they offer to the faithful a comforting

belief in divine authority over human affairs (Pollner 1989). As C.S. Lewis (1962) wrote,

> nor have I anything to offer my readers except my conviction that when pain is to be borne, a little courage helps more than much knowledge, a little human sympathy more than much courage, and the least tincture of the love of God more than all. (10)

Here too, religious involvement is multidimensional. In the language of social research, turning to religion is a kind of emotional coping strategy, one that is especially effective at reducing distress and restoring hope (Kaplan, Munroe-Blum & Blazer 1993) among the elderly (Koenig, George & Siegler 1988) and those experiencing illness or loss (Mattlin, Wethington & Kessler 1990). Religion legitimates the marginal human situations of illness and death by giving them a place in a single sacred reality (Berger 1967). Religious faith, exercised as private prayer, is a cognitive and emotional resource immediately accessible to the sick or disabled (Koenig 1993); the "social" support received from divine others (Pollner 1989) is similarly available, whatever the situation of need. Empirically, then, we should see data in which religious involvement is associated with poorer health cross-sectionally, and even over short followup periods (for examples, see Idler & Kasl 1992; King, Speck & Thomas 1994). In this view, the sign of the association is negative, and the causal order is reversed: health problems precede religious involvement.

In some extreme cases of the religious life, the strength of the negative association can even be intensified. A recent novel, *Mariette in Ecstasy*, is a fictional account of an intensely religious young girl who enters a convent and receives the *stigmata*, the wounds of Christ, in her hands, feet, and sides (Hansen 1991). This is a miraculous occurrence that was first experienced by St. Francis of Assisi, and subsequently by a small number of others whom the Catholic church has recognized and canonized. A negative association between religion and health would result from individuals turning to religion in times of illness, but an even more negative association would be seen when extremely devout individuals actually seek out physical suffering as a means of spiritual purification. Christians believe that not only Christ's death on the cross, but also his suffering there, atoned for their sins. For some, although not many perhaps, physical suffering may take on the religious meaning of identity with Christ; with the mortification of the flesh, the spirit is purified. These are extraordinary individuals, of course, "religious virtuosi," but the issue shows just how complicated things can get.

View 3: Religion and Self-Ratings of Health

A third way of thinking about religion and health hinges on the way health is defined and measured. It involves the self-ratings of health that people give as responses to the survey question, "Would you rate your health as excellent, good, fair, or poor?" Their responses, often considered too subjective to be of use in epidemiological studies, have been shown repeatedly to predict mortality over both short- and long-term periods of followup, even after adjusting for the respondent's medical condition (Mossey & Shapiro 1982; Schoenfeld et al. 1994).

While the meaning of these self-ratings of health remains uncertain, they apparently reflect a respondent's definition of health that encompasses more than just disease status, at least as it has been measured by surveys, medical records, or standard exams by a physician (Idler 1992). In fact, when asked about the meanings attached to their ratings ("When you answered the last question about your health, what did you think of?"), people offer answers citing their ability to work, their feelings of distress, their good or bad health habits, their own health in the past, or the health of others, in addition to their pain and present illness or injury (Groves, Fultz & Martin 1992). Students of the cognitive aspects of survey design have recently become very interested in the multiple meanings respondents assign to their answers to this apparently simple question (Feinberg, Loftus & Tanur 1985; Schechter 1994).

The breadth of the answers suggests that individuals are evaluating many areas of their life when assessing their health — including the health of others in comparison with their own, their emotional well-being, their social relationships, and their ability to perform valued activities in addition to their physical health. The relative weight placed on each of these areas and their patterns will differ from one individual to another. For example, one respondent in the present study, a 28-year-old Asian male with post-polio syndrome, answered that his health was excellent (in spite of the braces on his legs and the cane he walked with):

> My health is excellent. In general, I'm in that part of my life. I'm 28 years old. A lot of people help me get through my life before I go to surgery. I went through the surgery 14 times. Now I can walk. I can do better than before. I'm satisfied with my life that's for sure. Also my family helps me and my friends help me. You have to have other personalities to help you.

Thus in spite of having a debilitating chronic condition, this man emphasizes the support of his social network, his overall satisfaction with life, and his improvement from an earlier period in his life, and it allows him to evaluate what some might consider poor health as excellent.

One can see these evaluations of health as reflecting underlying concepts of the self; for some individuals, peak physical health may be an important aspect of their identity, for others it may be only incidental. The level of physical health necessary for this latter group to perform their social roles effectively or to gain satisfaction from valued activities may be minimal. Stephen Hawking, the physicist with severe multiple sclerosis, or Chuck Close, the painter who was partially paralyzed in midlife, would be extreme examples of people who have made major intellectual or artistic accomplishments despite severe physical debility.

We might very crudely break these areas of the self into two, and call them physical and nonphysical senses of self. Nonphysical senses of self would include cognitive or intellectual abilities, aesthetic interests in music, literature, or art, being a member of a profession, or having satisfying social relationships. The balance of one's physical and nonphysical senses of self might be stable throughout life, or it may shift as we age or experience illness. Charmaz (1991:259) writes that ill people who "define essential qualities of self as distinct from their bodies," especially those who are encouraged by others to do so, are

more likely to transcend the situation of their illness; they have "defined a valued self beyond a failing body." Development or possession of an intact nonphysical sense of self would then be associated with better global self-ratings of health even in the presence of chronic illness.

A religious or spiritual sense of self, the focus of this article, would certainly be among these aspects of identity. We might specifically predict that people who think of themselves as strongly religious or spiritual people would not find that identity threatened by illness or disability; indeed, the socially supportive attention of the religious group, and comfort received from prayer and ritual activities, could strengthen religious beliefs and reinforce such an identity. The sign of the association and the causal order in this third view are identical to the first: religious involvement precedes and is associated with better health. What differentiates the two is the meaning and measurement of the concept of health: self-ratings of health may be relatively positive, even in the presence of serious or chronic illness. There is some evidence for this association: one recent study shows religious practices to be positively related to an index containing two self-ratings of health (Ferraro & Albrecht-Jensen 1991).

Although the three views have been presented as incompatible with each other, their incompatibility is to some extent a heuristic simplification. Ultimately, all causal processes and directions of association could have a place in a complicated picture of processes that occur over time as people age, health events occur, and religious involvement declines or grows.

The hypotheses for this study are derived from the latter two ways of thinking about religion and health. First, with respect to the religious response to illness and suffering, we expect to see that respondents with poorer health or greater disability are more likely to say they have sought help from their religion. Second, with respect to religious and nonphysical senses of self, we expect to see that respondents with stronger religious senses of self-identity and better self-ratings of health will have a stronger nonphysical sense of self.

Method

SAMPLE

The participants in this study were randomly selected from clients of an urban rehabilitation clinic. The present study was part of a larger effort to better understand the meaning of self-ratings of health, with the eventual goal of discovering the processes by which these highly predictive self-ratings are generated. The strategy for the study was to focus on clients whose self-ratings were different from what might be expected from their observed physical health.

To identify such individuals, two instruments from the NHANES-I (National Center for Health Statistics 1977) were used to compare data on the same condition from two independent sources: a physician's examination of the client's musculoskeletal system and the patient's own self-report of pain in those same joints. The protocol for the physician's exam was the musculoskeletal exam portion of the first National Health and Nutrition Examination Survey (NHANES-I) (National Center for Health Statistics 1977). The examination

requires the physician to manipulate, flex, extend, bend, and rotate the client's joints, noting tenderness, swelling, deformity, limitation of motion, pain, and other symptoms of the knees, hips, back, neck, shoulders, elbows, wrists, ankles, feet, fingers, and toes. The second instrument, measuring the respondent's perceptions, was the *Arthritis Supplement* to the NHANES-I (National Center for Health Statistics 1977), a detailed set of questions about the individual's experience of pain, stiffness, and swelling in the knees, hips, back/neck, and other joints, which corresponds closely to the physician's examination.

Two hundred clients were screened with both instruments, representing a completion rate of 63.3%, to locate individuals who were reporting more pain than might be expected on the basis of their physical exam, and also those who reported less. These "pain" and "exam" scores could then be compared to the nationally representative NHANES-I sample from which the measures were taken.

Scores for the physician's evaluation (Exam) and the corresponding client reports of joint pain (Pain) were converted to Z-scores and the difference was taken, so that higher (positive) scores represented relatively greater physician findings, and lower (negative) scores represented greater reported pain. A similar difference variable was calculated for the entire original NHANES-I sample. Respondents in the clinic sample whose scores were one or more standard deviations away from the mean of the same variable for the NHANES-I sample were asked to participate in the psychosocial interview. Ironically, the screening was mostly unnecessary. In the NHANES-I sample, approximately 15% of the sample fell outside the one standard deviation line; in the clinic sample it was 88%. In other words, the great majority of the clients screened had pain reports that were different from what would be expected on the basis of their physician's exam. Because the clinic respondents' scores were compared to the NHANES-I distribution, not their own, this finding is *not* an artifact of the distribution of clinic scores. Most important, these scores were about evenly split between overreporters and underreporters.

Of the 176 clients who were screened into the second part of the study, 7 refused to participate further, 3 died or were hospitalized, and 20 could not be recontacted with the information they had given. This yielded 146 psychosocial interviews, or an 83% completion rate for the second part of the study.

MEASURES

Findings from the NHANES-I physician exam were summed for all sites.

For the *Arthritis Supplement*, if clients responded *no* to each of five screening questions, they were considered to have no musculoskeletal pain. As the emphasis in the present study was also on the interpretation of current states, respondents who reported pain of any duration that had not recurred in the last year were also considered to have no pain.

The health assessment instrument also includes the 38-item chronic conditions history of the NHANES-I Medical History Questionnaire (National Center for Health Statistics 1977), a body systems symptom review instrument — from which musculoskeletal symptoms were omitted for the purposes of this study (Cameron, Leventhal & Leventhal 1993); a 10-item physical abilities

battery used in the Rand Health Insurance Experiment (Brook et al. 1979) that asks the respondents about their ability to do simple activities (such as dressing and walking inside the house) and more difficult tasks (such as housecleaning and participating in sports), sociodemographic variables, health practices, and the single-item self-rating of health ("in general, would you say your health is excellent, very good, good, fair, or poor?"). This questionnaire was administered to the clients as soon as possible after the physician's examination, usually immediately.

The chronic-condition items were developed into two measures, one summarizing the respondent's chronic disease burden that puts them at risk for mortality and the other for chronic diseases that do not have a mortality risk but do cause disability. The measures were based on analyses of 16-year age-adjusted mortality risks by condition in data from the 1987 NHANES-I Epidemiologic Followup Study (NHEFS) (National Center for Health Statistics 1992).

The psychosocial interview took about one hour to complete. It combined frequently used scales and open-ended questions on a variety of subjects, including religion. Because self-ratings of health appear to reflect broad areas of emotional and social well-being and outlook on life (Adelman 1994; Maddox 1964), topics covered in the interview included social networks and support (Norbeck, Lindsey & Carrieri 1981), depression (Radloff 1977), neuroticism (Eysenck 1958), optimism (Scheier & Carver 1985), body consciousness (Miller, Murphy & Buss 1981), religious affiliation, activities, beliefs, and self-rated religiousness (Hoge 1972; McCready & Greeley 1976), and a set of items measuring the nonphysical sense of self. The last scale was developed specifically for this research, to elicit the importance the respondent placed on various aspects of nonphysical (social, intellectual, aesthetic, spiritual) identity. Thus they were asked, on a scale of 0 to 10, "to what extent do you think of yourself as a *musical* person? (Probe: If 10 is the most *musical* you could be, and 0 is not *musical* at all . . .?)" The 15-item scale appears in the Appendix to this article.

We used three measures of religious involvement. The first is a self-assessment of religiousness, based on the two asterisked items in the Appendix. The second is a measure of religious activities, including attendance at services and other religious activities, reading the Bible, knowing other people in the congregation, and watching religious programs on television. Finally, because of this article's concern with the issue of seeking help from religion for health problems, responses to an open-ended question were coded dichotomously for whether the respondent had received help or not. The respondent's religious affiliation was also recorded.

ANALYSIS

Cronbach's alphas were calculated for the appropriate scales: self-rated religiousness alpha = .63, religious activities alpha = .76, neuroticism alpha = .72, optimism alpha = .77, social body consciousness alpha = .67, depression alpha = .91, and nonphysical sense of self = .77.

Scale analysis, bivariate, logistic regression, and ordinary least-squares regression analyses were conducted for the two dependent variables: help from

religion and nonphysical sense of self. Multivariate analyses were given a hierarchical structure, with the religion variables entering the model first, to distinguish the dependent variable from other dimensions of religiosity. Sets of variables were then introduced one at a time; in the second step the key independent variables for health were added, and possible confounding demographic variables at the third step. Possible explanatory factors of health practices were added at the fourth step, social networks and support at a fifth step, and psychological characteristics and distress were added at the sixth step. Variables were included in the sets only if they had a significant bivariate association with the dependent variable in initial analyses; nonsignificant variables ($p<.10$) were dropped to reduce the number of variables in the model at any one time.

Results

Because this study adopted some of its instruments from those used with a nationally representative sample, the disability clinic sample can be compared with the larger sample (Idler 1993b). Respondents in the clinic sample were about evenly split between males and females, but the clinic respondents were much more likely to be nonwhite and unmarried, and they were somewhat more educated than the NHANES-I sample. The clinic sample had poorer self-rated health, more mortality-risk and disability-risk chronic conditions, poorer musculoskeletal status by physician's evaluations, and greater reports of pain.

Variables to be included in the analysis are described in Table 1. The two dependent variables, help from religion and nonphysical sense of self, are not at all correlated with each other.

Self-rated religiousness is highly correlated with receiving help from religion, as are race, education, and income. The other measures of religious involvement (activities and beliefs) do not appear because, while they were also highly correlated with receiving help from religion, they were too highly correlated with self-rated religiousness to be included together in a multivariate analysis. Nonphysical sense of self is not highly correlated with anything, although it has small associations with social support and neuroticism.

Table 2 shows the religious heterogeneity of the sample. Protestant respondents were more often nonwhite and younger than Jewish and other respondents. Roman Catholics were about evenly split between white and nonwhite and had the least education of the groups. Jewish respondents were more likely to be female, were all white, and had the oldest average age. Those with "other" or no religious affiliation were the most educated. Because of the demographic differences between the groups, and because of the similarities between Jewish and other/none respondents, two dummy variables, for Protestants and Catholics, were coded and included in all analyses. Jewish and other/none respondents were far less likely to report receiving help from religion. Nonphysical senses of self do not differ significantly among the groups.

The multivariate analysis begins in Table 3. The hypothesis whose testing is displayed in this table is that respondents with poorer health will be more likely to have sought help from religion. Variables included in the table are entered in

TABLE 1: Descriptive Statistics for Variables Used in Analysis, Disability Clinic Study

Variable	N	Minimum	Maximum	χ	Pearson r Help	Pearson r Nonp.
Help from religion	138	0 = Not mentioned	1 = Mentioned	0.6	1.00	.01
Nonphysical sense of self	143	23 = Low	111 = High	75.8	.01	1.00
Protestant	143	0 = Other	1 = Protestant	0.3	.24	.04
Catholic	143	0 = Other	1 = Catholic	0.3	.09	-.05
Self-rated religiousness	143	0 = Not at all religious	20 = Highly religious	12.1	.51	.17
Self-rated health	145	1 = Poor	5 = Excellent	2.7	-.04	.16
Disability	200	0 = None	20 = Severe	7.1	.19	.01
Chronic disease: Disability risk	200	0	1	0.5	-.18	.07
Cognitive status	200	1 = No impairment	3 = Impaired	1.2	.16	.06
Female	200	1 = Male	2 = Female	1.6	-.14	.05
Age	200	18	92	47.8	-.31	-.01
Race	200	1 = White	2 = Nonwhite	1.5	.57	.03
Income	188	$7,500	$80,000	$19,700	-.42	.07
Education	199	4 years	26 years	13.1	-.43	.11
Alcohol use	200	0 = Never	365 = Daily	47.4	-.27	.08
Inactivity	200	2 = Much exercise	6 = Little or no exercise	4.5	-.15	-.19
Social support: From respondent	144	9 = Little support	510 = Much support	178.4	.02	.21
Social support: To respondent	144	12 = Little support	487 = Much support	165.1	-.01	.23
Social network size	144	1 person	20 people	7.4	.02	.21
Frequency of social contacts, per year	145	18	4468	1,166.5	.07	.18
Depression symptoms	145	20 = Rarely	67 = Most of the time	36.5	.13	-.16
Neuroticism	142	0 = Low	6 = High	2.9	.14	-.21
Optimism	142	6 = Low	20 = High	14.7	.23	.19
Social body consciousness	144	0 = Low	12 = High	9.4	.21	.11
Physical body consciousness	144	0 = Low	16 = High	10.2	.16	-.08

* Correlations larger than .17 significant at $p < .05$

TABLE 2: Descriptive Statistics for Religious Affiliation Categories

Religious Preference	Percent	Percent Female[a]	Percent White[b]	Mean Age[c]	Mean Educ.[d]	Percent w/ Help from Religion[e]	Mean Nonphys. Sense of Self[a]
Protestant	31.5	55.6	20.0	43.2	12.7	31.2	76.9
Catholic	30.8	61.4	47.7	43.4	12.1	31.9	74.5
Jewish	17.5	80.0	100.0	65.1	14.4	16.7	73.6
Other/ none	20.3	62.1	51.7	53.0	15.1	20.3	78.1

[a] No significant differences between groups
[b] $\chi^2 = 41.3$, 3 df, $p < .0001$
[c] $F = 11.46$, 3 df, $p < .0001$
Bonferroni adjustment shows Jewish respondents older than all others.
[d] $F = 6.68$, 3 df, $p < .0003$. Bonferroni adjustment shows Roman Catholics with less education than all others, Other with more.
[e] $F = 7.3$, 3 df, $p < .0001$. Bonferroni adjustment shows Jewish respondents less likely than all others to receive help from religion.

sets from those that were significantly associated with receiving help from religion in Table 1. Tables 3 and 4 show standardized coefficients.

The analysis begins with model 1, considering just the religion variables. Respondents who think of themselves as strongly religious people are much more likely to say they have received help from religion, as are Protestants when compared with the Jewish/other/none respondents.

In model 2 we add the health measures that had a significant association with receiving help in Table 1, functional disability, cognitive status, and chronic disease disability risk. Only the disability measure is associated with receiving help, and this measure is retained throughout. In model 3 are added the sociodemographic factors; nonwhite respondents were much more likely to say they had received help, and lower levels of income and education were marginally significant, so all three are retained. Note that the significant effect of being Protestant disappears because of the strong effect of race.

In model 4 the health risks of alcohol use and inactivity are added; inactivity has a negative association with receiving help. In model 5 the psychological characteristics of neuroticism, optimism, body consciousness, and depression are added, but there are no significant relationships among them.

Thus, for view 2, the hypothesis is confirmed: the more disabled people are, the more likely they were to say they had sought help from religion. The direction of the association is that poorer health is associated with greater religiousness. This association is net of the respondents' ratings of their own religiousness, race, education, and activity levels.

TABLE 3: Regression of Help from Religion on Religion, Health Status and Psychological Factors, Disability Clinic Study

	Model 1	Model 2	Model 3	Model 4	Model 5
	B	B	B	B	B
Protestant	.293*	.287*	.049	.087	.131
Catholic	.233+	.170	-.069	-.060	-.071
Self-rated religiousness	.684***	.688***	.686***	.752***	.747***
Chronic disease:		-.169			
Disability risk					
Disability		.263*	.315+	.468*	.498*
Cognitive status		.101			
Female			-.031		
Race (nonwhite)			.590**	.591**	.669***
Income			-.291+	-.287	-.251
Education			-.355+	-.405+	-.490*
Age			-.146		
Alcohol use				-.194	
Inactivity				-.424*	-.397*
Neuroticism					-.035
Optimism					-.064
Body consciousness: Physical					-.272
Social					.021
Depression					.272
-2 Log likelihood χ^2	45.8	53.2	89.0	93.4	93.6
Degrees of freedom	3	6	9	9	13
N	138	138	129	129	128

* p < .05   ** p < .01   *** p < .001

In Table 4 we see the analysis of nonphysical sense of self. The direction of the hypothesized relationship with self-rated religiousness is the same as it was for help from religion; people with a stronger religious or spiritual sense of self should have a stronger overall nonphysical sense of self. The hypothesized direction for self-rated health, however, is different from the disability/help from religion association in the same way that view 2 is opposite to view 3: we expect people with a stronger nonphysical sense of self to have better self-rated health.

The analysis follows a similar strategy. In model 1, including just the religion variable self-rated religiousness is associated with nonphysical sense of self, but Protestants and Catholics show no differences from the Jewish/other/none respondents. These associations are not changed when self-rated health is added.

In the next column, the sociodemographic factor of education is added. It is marginally significant and is retained in further steps; as one would expect, higher levels of education correspond with greater nonphysical senses of self. Among the health practices, which originally included smoking, alcohol and illicit drug consumption, and inactivity (lack of exercise), only lack of exercise was associated at the bivariate level with either help from religion or nonphysical sense of self. Lack of exercise is entered in the fourth column here, and it shows no significant association with nonphysical sense of self.

Several aspects of social networks and support were associated at the bivariate level with nonphysical sense of self, but none of the associations are significant when tested in the multivariate model.

Last, we look at some psychological characteristics, which had significant bivariate associations. None of them, however, is significant in the multivariate analysis.

Thus the most parsimonious model containing all significant effects is model 3, in which the nonphysical sense of self is associated with higher levels of religiousness, better self-ratings of health, and to some extent, higher levels of education, as was hypothesized.

Discussion

This study has many limitations; at best it can be considered illustrative and preliminary. The sample is small and highly selected for individuals with health problems. Moreover, it is a cross-sectional study in a field in which the causal directions are especially problematic. At most it can suggest directions for future studies of religion and health.

With these limitations in mind, however, the seemingly paradoxical hypotheses for this study were both supported, and they show that religion does have a complicated relationship with health. The first hypothesis was that respondents with greater disability would be more likely to say they had sought help for their health problems from their religion. This yes/no response was coded from what was actually an open-ended question in the interview: "Do your religious beliefs help you at all with your health problems? Can you tell me in what way, or give me any specific examples of help that you received?"

TABLE 4: Regression of Nonphysical Sense of Self on Religion, Health Status, and Psychological Factors Disability Clinic Study

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Variables	B	B	B	B	B	B
Self-rated religiousness	.184*	.203*	.226*	.220*	.225*	.219*
Protestant	-.042	-.111	-.055	-.071	-.086	-.071
Catholic	-.075	-.122	-.053	-.075	-.031	-.041
Self-rated health		.198*	.181*	.149+	.146+	.132
Education			.150+	.143	.173+	.130
Inactivity				-.136		
Social support: To respondent					.202	
From respondent					-.067	
Social network size					.009	
Frequency of social contacts					.113	
Neuroticism						-.158
Optimism						.081
Depression						-.003
Adjusted R^2	.01	.04	.05	.06	.07	.06
N	143	142	141	141	140	140

+ p < .10 * p < .05 ** p < .01 *** p < .001

There were many interesting answers to this question; 62% of the sample gave a positive response. In fact, the respondents' stories illuminate both findings of the study; they especially provide insight into the meaning of identities based on spiritual or religious criteria.

SPIRITUAL AWAKENINGS

First there are accounts from respondents who felt that help from God or their religion changed everything in their lives. Perhaps the most dramatic stories were of religious awakenings that took place directly as a result of a sudden illness or injury.

> I don't know . . . I felt so spiritual, you know [after the stroke]. It was such a — you know — a beautiful experience. I get like that every now and then. . . . I got this relaxed feeling and feel so good. I don't think you could feel that good all the time. We couldn't walk around feeling like that daily. One thing, I stopped worrying as much as I used to, and I been able to let go of certain things that I couldn't let go of before. And, you know, I used to sit and cry because I was so happy. Yeah, it's a great way to feel. . . . I learned to be by myself. I get lonely sometimes . . . but I can deal with it more now. So in a sense I look at my stroke as a blessing. . . . I think . . . through this stroke I started to feel more spiritual.

These are the words of a 51-year-old black Protestant woman who had been a typist before her stroke. She is still partially paralyzed and speaks elsewhere of looking down at her hands, which she cannot move. Rather than regret, however, at her loss of her skill and ability to earn a living, she thinks of those who have had a stroke and died, and she feels lucky to be alive. More than that even, she feels that her life has been enriched by the experience, so much that she used to "sit and cry because I was so happy" and actually speaks of her stroke as a blessing. There were other such stories. Another was from a 35-year-old black Protestant woman who had also been a secretary and had also had a stroke:

> At the time before my stroke I had no experience with my God. With my stroke I learned from him. It happened with faith. My pastor and my friends talked to me about religion. I was not very serious about religion. I take God seriously, deeply. People in my church pray for me and visit me all the time. They call me and give me support.

Both these women identified their illnesses as real turning points in their religious lives, although they imply that they had had some connection with religion ("my pastor and my friends talked to me") before the event. One man, also a Protestant, 52 years old, white, a fabric designer who was paralyzed after a mugging in which he was stabbed in the back, explained,

> When your life drastically changes I think if you had any religious training or background you tend to resort or turn to it or think of it or find it more enhancing or helpful. Due to my injury I've become more religious.

Whether they are awakenings or reawakenings isn't really important. What is important is that these respondents saw a growth in their spiritual lives that coincided with, and came about as a direct result of, their shrinking physical capacity. A loss in one area of life became a gain in another. It is especially clear

when the respondents had some degree of paralysis, an unambiguous loss of physical ability. It is a causal chain these people are telling us about, one in which they consciously make the direct connection between the helping functions of religion and the nonphysical, religious or spiritual sense of self. One more example is from a 25-year-old Protestant Hispanic male who had injuries that resulted in memory loss and speech problems:

> Through my injuries I'm so depressed. When I go to church, I feel so much better. It relieves me. A few times I've been to church, I "caught the Spirit." And after that I felt so much better.

DISCOVERING PURPOSE IN LIFE

There were some other respondents who spoke about religious awakening in a different way, not as a feeling of spirituality, but as a conviction that God had a purpose for them. A number of these people's illnesses or injuries had brought them close to death, and the very fact of their survival took on a religious meaning, causing them to see their lives in a new way. One man was a 42-year-old Russian immigrant, a bus driver, who had fallen five flights of stairs and injured his pelvis and head and had nerve damage (his religious affiliation was "other," for Russian Orthodox):

> When I had my accident, I had spent time in a coma for three weeks. And finally I survived, and was on my way to recover, and I think that if God saved my life, he had a purpose. So I believe him, and I'm not going marching down.

Another man, a 40-year-old Hispanic, Roman Catholic man whose legs had both been amputated after a train accident said simply,

> I thank God. There is a purpose to why I'm still alive.

HEALING, STRENGTH, AND COMFORT

In many other cases, there was, in addition to gratitude for being alive, also gratitude for the recovery they had experienced, which they often framed in terms of religious healing. Perhaps the single most dramatic example came from a pretty young Protestant Hispanic woman who had suffered severe head injuries in a car accident. She limped slowly into the interview room on crutches and with braces on both legs. She told me she had been in a coma for three months:

> [The doctors] told me I would never walk again. [What helps me is] trusting in God, knowing when it's his time. I ask for his help and he always helps me. He is there caring for me, he's dependable.

The interviewer said, "And you just walked in here." And she beamed. Another woman (Roman Catholic) who asked God directly for healing and felt she received it was 63 years old and had immigrated from the Virgin Islands. She had had a stroke, was using a walker, and still had partial paralysis in one arm:

> When I was in the hospital, when I woke up and I couldn't feel nothing, I figured that I would be an invalid. I would be in a wheelchair, and people would have to do things for me. And then I started praying, asking Jesus to help me, give me strength, that I could do

the things, and have nobody wait on me. My prayer was answered. I've come a long way, and I have to give thanks to almighty God. I have to thank my God every day.

Another example is from a young black Protestant woman who fell four stories from her apartment in a fire:

I think that believing in God . . . a lot of people thought I was going to die. I'm making it. Even the doctor told me that I have nine lives because they almost gave me up. With God's help I'm walking around. I didn't move around for eight months. I was on my back. I couldn't do anything for myself. Now with the help of God I'm doing all these things.

There were several other respondents who reflected directly on the relationship between the healing that comes from medicine and the healing that comes from faith. For example,

When medicine can't help, God does. He's the doctor of all doctors.

A 45-year-old black auto repairman, a Protestant, who had had a stroke said

[My religious beliefs] helped 100%. When I was going to the church I couldn't walk. I prayed a lot and I noticed I made an improvement. I believe it was partly because of the prayer. Again, it's the Lord who made us. He's the one who knows about us. Doctors are directed by the Lord. The physician's knowledge comes from God. I'm sick with a stroke and the doctor might say I'm not going to walk or I may not be able to live again. They are directed by the Lord. I think all my improvement comes from God.

Respondents conceived of "help" in many different ways. Short of attributing any healing to their faith, many respondents at least felt they gained peace, encouragement, strength, and comfort. Here is a sampling of responses:

I try to think of the hereafter. That gives a little optimism I think. Just the occasional dropping into church gives some encouragement. (69-year-old white Roman Catholic woman, former secretary with back pain)

The doctor said if the car hit me by another inch I would have been crippled. God helped me with that. In my depression I've asked him to give me the strength and to lift my burdens. (43-year-old black Protestant man, computer supervisor)

It's helped me overcome. I didn't take public transportation, but now I take it. God makes me feel brave. I feel more independent. (32-year-old Asian woman, Protestant, data entry worker with post-polio syndrome)

If I'm in pain whether it be physical or mental, I pray for strength and guidance and I feel that I've gotten it. (36-year-old black woman, Protestant, clerical worker with epilepsy)

Yes, it's comforting knowing that God is there to help me deal with pain and everything else in life. (40-year-old black man, Roman Catholic, office associate with vascular problem)

FINDING A MEANING IN SUFFERING

There was also a theme of enduring suffering that several respondents developed. Some framed their experiences in terms of being tested by God, and believing that their faith would be rewarded in the future. These were respondents who could point neither to tangible physical healing nor to intangible emotional comfort, but who believed that their present suffering must have some religious meaning. Here are some examples:

I am attending a group that believes in reincarnation. We discuss what the soul is, what a person is. I believe life is a learning experience. I believe the soul goes through many lifetimes. I believe that my disability is something that I have to go through to sort of get back on the right path of evolution. We are all evolving into something perfect. People who choose difficult lives choose them to learn from them. (43-year-old Jewish female, newsstand worker with arthritis of the knee)

Job suffered and endured due to his faith in God. His reward then was twice as much. Endurance is the most important thing and faith that he does exist and that if we do his will the rest comes. (49-year-old Hispanic Protestant woman, saleslady suffering pain and depression)

My friend, when she was sick, she never thought that she was going to die. She thought that maybe God was thinking about her and wanted to see if she really had faith in him. (22-year-old Hispanic woman with lupus. She spoke so often during the interview of "her friend" that one wonders if she wasn't really talking about herself)

Because they say God has no respect for individual persons. Anyone can get sick because tomorrow isn't promised to anyone. Just thank God for one more day. Anything can happen. It's scary, you have people, family, kids. The environment around us, you tend to worry about families. I feel prayer helps. Reading in the Bible that Job had to go through sickness even though he was a perfect and upright man and everything. God blessed him more and gave him a new family. (24-year-old black Protestant woman with multiple sclerosis)

These accounts are among the most thoughtful and subtle that were given heard. They show a poignant level of reflection, a sense of prayers prayed but not answered, a struggle, a real groping for meaning that does not come easily. Their identification with the biblical figure of Job, especially, shows the unresolved nature of their suffering.

Like Job, some respondents were also angry with God for allowing evil in the world, or cruelty, or neglect. One 76-year-old actress who had no religious affiliation and had suffered a stroke was being interviewed during the Los Angeles riots. She said

You could make a case for that [that religious beliefs help people with their pain or health problems]. I could make a case. I sometimes read things and I become completely swayed by what I read and then I think, well, I get cynical about it. I can't answer that question. I wish I had that connection, I really do. I've prayed in my life. I prayed a lot when I was little. When something happens like it's happening now in Los Angeles, how can you think that this God in heaven cares about you? It's stupid.

Another woman, a 53-year-old Hispanic Roman Catholic who had had an aneurysm, gave a classic account of the problem of the existence of evil in the world:

I should be more closer to God, you know. I can't believe that I'm like this. God is supposed to be kind, love, you know. People say that, I can't. If God is supposed to be so kind, so helpful, he would cure me if I was sick. Quote the Bible, blind people, crippled people heal ye . . . . from death. How come he wants people to go hungry? To be paralyzed, even kill each other? God cannot be giving me this cause I refuse, coming from the Devil, the Devil is mean. But that's different stuff. Or maybe I wicked in my own self and I don't know it, and I deserve this. We've been punished ever since Adam and Eve. So maybe I would say well then why should I worry about it. They did it and that's how we have to pay for it. Maybe I should take this penalty, but I can't. Because I should not pay for what they did.

Another woman blamed God more indirectly, for the failings of the people around her to help her as she thought they should. This was a 36-year-old Hispanic woman, a Roman Catholic:

God helps me in a lot of ways. But he put people in this world to help and I don't see anyone helping me in the things I really need. I need physical therapy, understanding, and to treat me like a normal person. I'm just like everybody else. I'm not getting the help I need.

THE SPIRITUAL, NONPHYSICAL SELF

The argument in this article has been that one of the benefits of religious involvement will be in refocusing the respondent on aspects of the self to which a painful, or nonfunctioning, or unattractive physical body is irrelevant. The inner, spiritual self can be beautiful and whole even when the material, physical body is broken or diseased. As these are important teachings of the Christian religion, it is not surprising that the responses above come almost exclusively from Protestants and Roman Catholics. But there was another important group of responses from people who were Jewish or identified their religious affiliation as other or none, and who also regarded their spiritual self as very important. One woman, a 37-year-old writer with Guillan-Barré syndrome said,

Yes [her religious beliefs helped her] because of the meditation. You lay down and localize where your pain is. From there you take the pain and you have a spirit at the end of the bed that receives the pain. You watch the pain go off. You go over every step of relaxation until the pain goes away. I've only been going there for two months. . . . there is a Science of Mind Center that I go to on the weekends. They have a healing circle and prosperity groups that focus on living a happy and healthy life-style without medication. There are practitioners there and a reverend. There is a religious aspect to it.

Another woman, a 48-year-old Jewish musician who had Ehlers-Danlos syndrome, a hereditary disease of the connective tissue, said

My inner values, my nonmaterial world values, are very helpful with my pain. I don't call it religion. I work with my body energy and my emotional energy. I use my visualizing techniques to manipulate my mind. To me they are spiritual exercises, choosing life as opposed to being struck down by my situation.

Finally, here are some passages from the very lengthy response of a 65-year-old European immigrant; she had had breast cancer some years before and was currently seeking treatment for arthritis. She was a professor and identified her religious affiliation as "other":

As I come close to my exit, I come to what Erikson calls "the closing of the circle." I was inspired by Herbert Benson and Norman Cousins. It is the state of being you are in, I am at peace with myself — our bodies can fight cancer themselves. I pat my breast, I say "Tootsie, we have to coexist. If I am cremated you are too." If it is not to be, God whoever you are, okay then please let me stay a little longer, and then go as fast as possible. I don't think suffering is noble. . . . I'm trying to explain what is beyond language. I'm smarter than this, it is liberating, I can go into any of them [churches] and feel at home. I don't live in fear — I am in resolution in my unconscious.

This woman was an intellectual, well-read and thoughtful, a professional who told me with great pride about students she had trained; and she reveals a remarkable separation between her physical self and her "real" nonphysical self in imagining conversations with her own body parts. She also was a lonely person, with no family, and she had recently lost her best friend to a recurrence of breast cancer; her self-rating of health was "fair."

Religion is a complicated thing. People seek help from it when they are in trouble, especially when they have been brought up with a traditional faith, and especially when that religious tradition speaks to their situation. So the data show, quite straightforwardly, that people with greater disability were more likely to say they had sought help from religion. The respondents' own words, which I have used liberally, seem to convey a consistent theme that people turn to religion in need, and that one of the ways religion helps them is by letting them "rise above" their problems, so to speak, by putting them in a context in which the physical body does not matter that much.

Well, I know that some people if they believe in God and if they believe in prayer would pray to be relieved of their pain. But I always could see others that were worse off than me so when I went to pray, I couldn't pray for myself when the guy next door was so much worse so I wound up praying for him and it helps to take my mind off my own problems. (54-year old Roman Catholic woman who had had a stroke)

In these voices speaking of religion and health there are themes of new insight, of the relevance of religious beliefs to some of the most challenging and threatening of human experiences, and of the growth and development of the self. Above all, they speak of the authenticity of these experiences. Self-ratings of health, another expression of subjective experience, are being taken more and more seriously by researchers because of their apparent ability to communicate vital health information. Their relationship with eventual health outcomes should also underscore for us the validity and importance of the patient/respondent's point of view, a view some have argued is increasingly absent in the historical development of Western medicine (Reiser 1993).

The coherence of these voices belies the contradictions in the relationship between religion and health presented at the beginning of this article. People say they do turn to religion in crises of illness, and religion also seems to be associated with more positive views of health when the spiritual self becomes more prominent. Ultimately, however, longitudinal data are necessary to establish causal order in these processes that appear contradictory at the cross section.

APPENDIX: Sense of Self Scale

Here are some ways in which people think about themselves. For these questions, I would like you to think about your whole life, not just the way you are today.

On a scale of 0 to 10, tell me to what extent you think of yourself as:

1. a musical person?
2. a reader?
3. an artistic person?
4. a social person?
5. an intellectual?
6. a family person?
7. a conversationalist?
8. a writer?
9. a spiritual person?*
10. someone who is good with numbers?
11. someone who is good at solving problems or puzzles?
12. a scientific person?
13. a professional?
14. someone with a sense of humor?
15. a religious person?*

* Treated as a separate scale for the purpose of this study.

References

Adelman, Pamela K. 1994. "Multiple Roles and Physical Health among Older Adults." *Research on Aging* 16:142-66.

Berger, Peter L. 1967. *The Sacred Canopy: Elements of a Sociological Theory of Religion*. Doubleday.

Brook, Robert H., John E. Ware, Allyson Davies-Avery, Anita L. Stewart, Cathy A. Donald, William H. Rogers, Kathleen N. Williams, and Shawn A. Johnston. 1979. "Overview of Adult Health Status Measures Fielded in Rand's Health Insurance Study." *Medical Care* 17 (Supplement):1-131.

Cameron, Linda, Elaine Leventhal, and Howard Leventhal. 1993. "Symptom Representations and Affect as Determinants of Care-Seeking in a Community-Dwelling, Adult Sample Population." *Health Psychology* 12:171-79.

Charmaz, Kathy. 1991. *Good Days, Bad Days: The Self in Chronic Illness and Time*. Rutgers University Press.

Durkheim, Émile. [1915] 1965. *Elementary Forms of the Religious Life*. Translated by Joseph W. Swain. Free Press.

_____. [1897] 1951. *Suicide*. Translated by John A. Spaulding and George Simpson. Free Press.

Eysenck, H.J. 1958. "A Short Questionnaire for the Measurement of Two Dimensions of Personality." *Journal of Applied Psychology* 42:14-17.

Feinberg, Stephen E., Elizabeth F. Loftus, and Judith M. Tanur. 1985. "Cognitive Aspects of Health Survey Methodology: An Overview." *Milbank Memorial Fund Quarterly* 63:547-64.

Ferraro, Kenneth F., and Cynthia M. Albrecht-Jensen. 1991. "Does Religion Influence Adult Health?" *Journal for the Scientific Study of Religion* 30:193-202.

Groves, Robert M., Nancy H. Fultz, and Elizabeth Martin. 1992. "Direct Questioning about Comprehension in a Survey Setting." Pp. 49-61 in *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*, edited by Judith M. Tanur. Russell Sage.

Hansen, Ron. 1991. *Mariette in Ecstasy*. Harper Collins.

Hoge, Dean R. 1972. "A Validated Intrinsic Religious Motivation Scale." *Journal for the Scientific Study of Religion* 11:369-76.

Idler, Ellen L. 1987. "Religious Involvement and the Health of the Elderly: Some Hypotheses and an Initial Test." *Social Forces* 66:226-38.

____. 1992. "Self-Assessed Health and Mortality: A Review of Studies." *International Review of Health Psychology* 1:33-54.

____. 1993a. *Cohesiveness and Coherence: Religion and the Health of the Elderly*. Garland.

____. 1993b. "Perceptions of Pain and Perceptions of Health." *Motivation and Emotion* 17:205-24.

Idler, Ellen L., and Ronald J. Angel. 1990. "Age, Chronic Pain, and Subjective Assessments of Health." *Advances in Medical Sociology* 1:131-52.

Idler, Ellen L., and Stanislav V. Kasl. 1991. "Health Perceptions and Survival: Do Global Evaluations of Health Status Really Predict Mortality?" *Journal of Gerontology: Social Sciences* 46:S55-65.

____. 1992. "Religion, Disability, Depression, and the Timing of Death." *American Journal of Sociology* 97:1052-79.

Kaplan, Berton H., Heather Munroe-Blum, and Dan G. Blazer. 1993. "Religion, Health, and Forgiveness." Pp. 52-77 in *Religion in Aging and Health*, edited by Jeffrey Levin. Sage.

King, Michael, Peter Speck, and Angela Thomas. 1994. "Spiritual and Religious Beliefs in Acute Illness — Is This a Feasible Area for Study?" *Social Science and Medicine* 38:631-36.

Koenig, Harold G. 1993. "Religion and Hope for the Disabled Elder." Pp. 18-51 in *Religion in Aging and Health*, edited by Jeffrey Levin. Sage.

Koenig, Harold G., Linda K. George, and Ilene C. Siegler. 1988. "The Use of Religion and Other Emotion-Regulating Coping Strategies among Older Adults." *The Gerontologist* 28:303-10.

Levin, Jeffrey S., and P.L. Schiller. 1987. "Is There a Religious Factor in Health?" *Journal of Religion and Health* 26:9-36.

Lewis, C.S. 1962. *The Problem of Pain*. Macmillan.

Maddox, George L. 1964. "Self-Assessment of Health: A Longitudinal Study of Selected Elderly Subjects." *Journal of Chronic Disease* 17:449-60.

McCready, William C., and Andrew Greeley. 1976. *The Ultimate Values of the American Population*. Sage.

Mattlin, Jay A., Elaine Wethington, and Ronald Kessler. 1990. "Situational Determinants of Coping and Coping Effectiveness." *Journal of Health and Social Behavior* 31:103-22.

Miller, Lynn C., Richard Murphy, and Arnold H. Buss. 1981. "Consciousness of Body: Private and Public." *Journal of Personality and Social Psychology* 41:397-406.

Mossey, Jana, and Evelyn Shapiro. 1982. "Self-rated Health: A Predictor of Mortality among the Elderly." *American Journal of Public Health* 72:800-808.

National Center for Health Statistics. 1977. *Plan and Operation of the Health and Nutrition Examination Survey, United States, 1971-1973*. (DHHS Publication no. PHS 79-1310). U.S. Public Health Service.

____. 1992. *Plan and Operation of the NHANES-I Epidemiologic Followup Study, 1987*. (DHHS Publication no. PHS 92-1303). U.S. Public Health Service.

Norbeck, Jane S., Ada M. Lindsey, and Virginia L. Carrieri. 1981. "The Development of an Instrument to Measure Social Support." *Nursing Research* 30:264-69.

Pollner, Melvin. 1989. "Divine Relations, Social Relations, and Well-Being." *Journal of Health and Social Behavior* 30:92-104.

Radloff, Lenore S. 1977. "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population." *Applied Psychological Measurement* 1:385-401.

Reiser, Stanley Joel. 1993. "The Era of the Patient: Using the Experience of Illness in Shaping the Missions of Health Care." *Journal of the American Medical Association* 269:1012-17.

Schechter, Susan (ed.). 1994. *Proceedings of the 1993 NCHS Cognitive Aspects of Self-Reported Health Status.* U.S. Department of Health and Human Services, Public Health Service, CDC, National Center for Health Statistics.

Scheier, Michael F., and Charles S. Carver. 1985. "Optimism, Coping, and Health: Assessment and Implications of Generalized Outcome Expectancies." *Health Psychology* 4:219-47.

Schoenfeld, David E., Lynda C. Malmrose, Dan G. Blazer, Deborah T. Gold, and Teresa E. Seeman. 1994. "Self-Rated Health and Mortality in the High-Functioning Elderly — A Closer Look at Healthy Individuals: MacArthur Field Study of Successful Aging." *Journal of Gerontology: Medical Sciences* 49:M109-M115.

Religious Resources and Church Growth*

LAURENCE R. IANNACCONE, *Santa Clara University*
DANIEL V.A. OLSON, *Indiana University-South Bend*
RODNEY STARK, *University of Washington*

Abstract

This article models the growth of religious organizations as a "product" derived from inputs of time and money. Using measures of church attendance and contributions as proxies for time and money resources, we predict membership growth at the levels of both individual congregations and entire denominations. Our data also highlight the tremendous variation in rates of resource mobilization across different denominations. In each of the past two decades, the decline of liberal/mainline denominations and the growth of their more strict/conservative counterparts can be modeled largely in terms of the radically different amounts of time and money that they demand and receive from members.

Why do some religious organizations grow rapidly while others lose members? Or, to address the primary historic trend in American religion, why have theologically conservative denominations prospered while the more liberal, "mainline" bodies declined?

More than 20 years ago, Dean Kelley ([1972] 1986) proposed an answer to this question. His now-classic study, *Why Conservative Churches Are Growing*, identified organizational "strictness" as the distinguishing feature of strong churches and thus traced the success of conservative denominations to their demands for complete loyalty, unwavering belief, and rigid adherence to a distinctive lifestyle. Recent work has used rational choice theory to formalize parts of Kelley's thesis and to demonstrate how strictness leads to organizational strength (Finke & Stark 1992; Iannaccone 1992, 1994). In particular, it has shown that strictness can function to screen out free riders, thereby raising average levels of commitment and participation within an organization. Like Kelley's book, however, the more recent work gives limited attention to the link between commitment and growth.

This essay tackles the subject of growth more directly. Building from the concepts of religious *resources* and religious *production*, it develops a theoretical

* *The first author's work on this article was supported in part by grants from the Lilly Endowment (#920370) and the University of Santa Clara. David Caddell provided valuable assistance collating data from the U.C.C. Church Membership Inventory. Direct correspondence to Laurence Iannaccone, Department of Economics, Santa Clara, CA 95053.*

model in which "inputs" of time and money combine to "produce" new members. Growth statistics, both within and across denominations, demonstrate the empirical power of this approach. Although growth remains the central concern, we also examine the weaknesses of traditional measures of religious participation, offer suggestions for improvement, and demonstrate how our approach can enhance research on religious markets and denominational differences.

Our arguments owe an obvious debt to resource mobilization scholars, who repeatedly have emphasized that "study of the aggregation of resources (money and labor) is crucial to an understanding of social movement activity" (McCarthy & Zald 1977: 1216; see also McAdam, McCarthy & Zald 1988; McCarthy & Wolfson 1993). It is, however, from their general perspective rather than their specific propositions that we have borrowed most heavily. Compared to social movement organizations, mainstream denominations interact less with external constituencies and employ a much narrower range of resource-mobilizing strategies and technologies. Most of their energies go toward meeting the needs of a relatively small circle of members, who are in the language of McCarthy and Zald (1977:1215) the denominations' principal (goal accepting) "adherents," (resource providing) "constituents," and (benefit receiving) "beneficiaries."[1]

Modeling Growth

A religious organization cannot survive, much less grow, unless it obtains sufficient resources from the environment. Congregations need places to meet and people to lead them; church structures cost money to build and maintain; outreach, evangelism, and community service demands time and money; the activities of ministers, music directors, Sunday school teachers, and janitors never come free, even when supplied by volunteers[2] — a great many other examples could be cited.

Growth can occur only if there are *surplus* resources, such as time and money beyond the minimum required to maintain current operations and to compensate for depreciation in physical facilities and in membership lost to death or departure. Surplus funds permit a church to hire additional staff to design and direct programs of outreach that meet the social, physical, and spiritual needs of potential new members. Income spent on new or enhanced programs can also provide more and better services for current members and thereby increase retention. Assets used to expand and improve meeting facilities can help attract and accommodate additional members. Donations of members' time prove no less important than donations of money. People are naturally drawn to churches with an energetic membership ready to help with potlucks, teach Sunday school classes, serve on committees, organize worship services, sing in the choir, assist in soup kitchens and community outreach, and help other members with special needs. All these activities require volunteered time and effort, and all help to attract and retain members.

To avoid needless detail, it helps to lump resources under two broad headings, labor and capital, or their rough equivalents, time and money. People's time and effort are combined with purchased inputs (both physical

goods and hired services) to produce valued commodities. It follows that growth requires "inputs" of time and money and that greater inputs tend to facilitate growth and inhibit decline (other things remaining equal). Input requirements are not rigid, however, since there are many ways to accomplish almost any outcome, including growth. One congregation may rely heavily upon paid professionals, whereas another may get by with lay preachers and other volunteers. Religious organizations can thus *substitute* between time and money, although technological and normative constraints limit the degree of input substitutability (Oliver & Marwell 1992).

Standard economic production functions capture these features in a single symbolic expression:

$$G = F(L,K)$$

where G represents growth, L represents labor, K represents capital, and the mathematical function F specifies the relationship between inputs and outputs. Roughly speaking, this translates to

rowth = ti e + one

Stated more precisely, the expression asserts that growth is an increasing function of two substitutable inputs, labor and capital.[3]

The model abstracts from numerous details so as to keep the analysis simple and thus suitable for generating testable propositions. This is not to deny the possibility of other growth-related factors, such as the religion's content, the organization's commitment to growth and evangelism (Greer 1993; Hadaway 1993; Olson 1993; Royle 1993), the members' socioeconomic attributes, the presence of competing organizations (Finke & Stark 1992), and trends in the economic, social, and demographic environment (Hadaway & Roozen 1993; Hoge & Roozen 1979a). But clearly one is hard pressed to fully define, much less formally analyze and empirically test, a model of the form

growth = time + money + content + goals + member attributes
+ competition + trends + social environment + ...

The proposed model zeros in on two fundamental factors — time and money.[4] With it, we hope to demonstrate that the principles of resource mobilization provide an important, but often overlooked, step toward understanding church growth.[5]

At the national level, where environmental factors tend to randomize out and affect all denominations more or less equally, we expect resource variables to be among the most important determinants of denominational growth. At the local level, where social, economic, and demographic factors differ greatly from one community to the next, we expect that congregational growth will depend on a variety of factors but that resources will still prove important.

Measuring Resources

If growth depends on inputs of labor and capital, then one must attend to their measurement. Since virtually all the resources at the disposal of a religious organization originate with its members, we can obtain a fairly complete accounting simply by measuring the time and money that members devote to

religious activity. Ideally, one should also measure external support — from government agencies or foreign organizations — and the additional resources derived from financial investments, church-owned properties, and church-run businesses.

Unfortunately, the traditional social scientific measures of religious commitment and participation were not designed to provide this information. Consider, for example, the standard question asked in Gallup surveys for the past fifty years: "Did you, yourself, happen to attend church in the last seven days?" Using this item to assess the quantity and quality of a church's religious labor force is a bit like assessing hours of employment by asking people if they worked any time during the last week. We will know who worked, but not how much. The Gallup question provides no way to distinguish between minimally productive bench warmers and dedicated church servants.

To clarify the problem, consider the members who attend a worship service every week but never spend additional time in church or on church-related activities. It consumes substantial resources merely to produce a weekly worship service — a church building must be purchased (or rented) and maintained; someone must be paid or persuaded to conduct the service; others must help to provide music, plan the service, act as ushers, and so on. As long as the attenders contribute time and money equal to the total costs of providing the service, the organization breaks even. But if contributions fall below total costs, as can easily happen if most members do no more than attend, the organization runs a deficit on the very people that survey analysts routinely classify as "committed" members.

Members who show up only for worship services — even those who do so regularly — provide their church with little or no surplus labor. Such people scarcely count as members of the church's labor force, except insofar as their presence, singing, testimonials, and prayers enhance the worship experience of those around them. A church's survival thus hinges upon the number of members who devote time *above and beyond* standard worship services. Getting core members to show up for planning meetings, work days, choir practice, fellowship groups, and weekday Bible studies may prove much more important than getting casual members to attend more regularly.

This fact highlights a major shortcoming in the relatively detailed measure of church attendance used in the General Social Survey. Though it provides no fewer than ten response categories to the question "How often do you attend religious services?" only one of these categories describes more-than-weekly involvement. Active members must therefore choose between just two responses: "every week" and "several times a week." From the resource-mobilization perspective, this question needs many more high-end categories, not to mention additional questions on the hours and content of the individual's religious activities.

Despite these failings, the GSS data suffice to demonstrate how unevenly surplus labor varies across denominations. Figure 1 shows the percentage of people in various denominations who claim to attend religious services "several times a week."[6] (When working with national survey data, one must note that denominational affiliation reflects the respondents' self-identification rather than their official membership standing.) Within the "mainline" denominations

(Episcopal, Disciples, Lutheran, Methodist, Presbyterian, UCC, and Catholics) the fraction rarely exceeds 4%. In contrast, Southern Baptists weigh in at 12% and the so-called sects (Adventists, Jehovah's Witnesses, Mormons, and numerous fundamentalist/evangelical or pentecostal/charismatic groups) routinely score over 20%.[7] Among Jehovah's Witnesses, "several times a week" attenders make up a whopping 44% of the total.

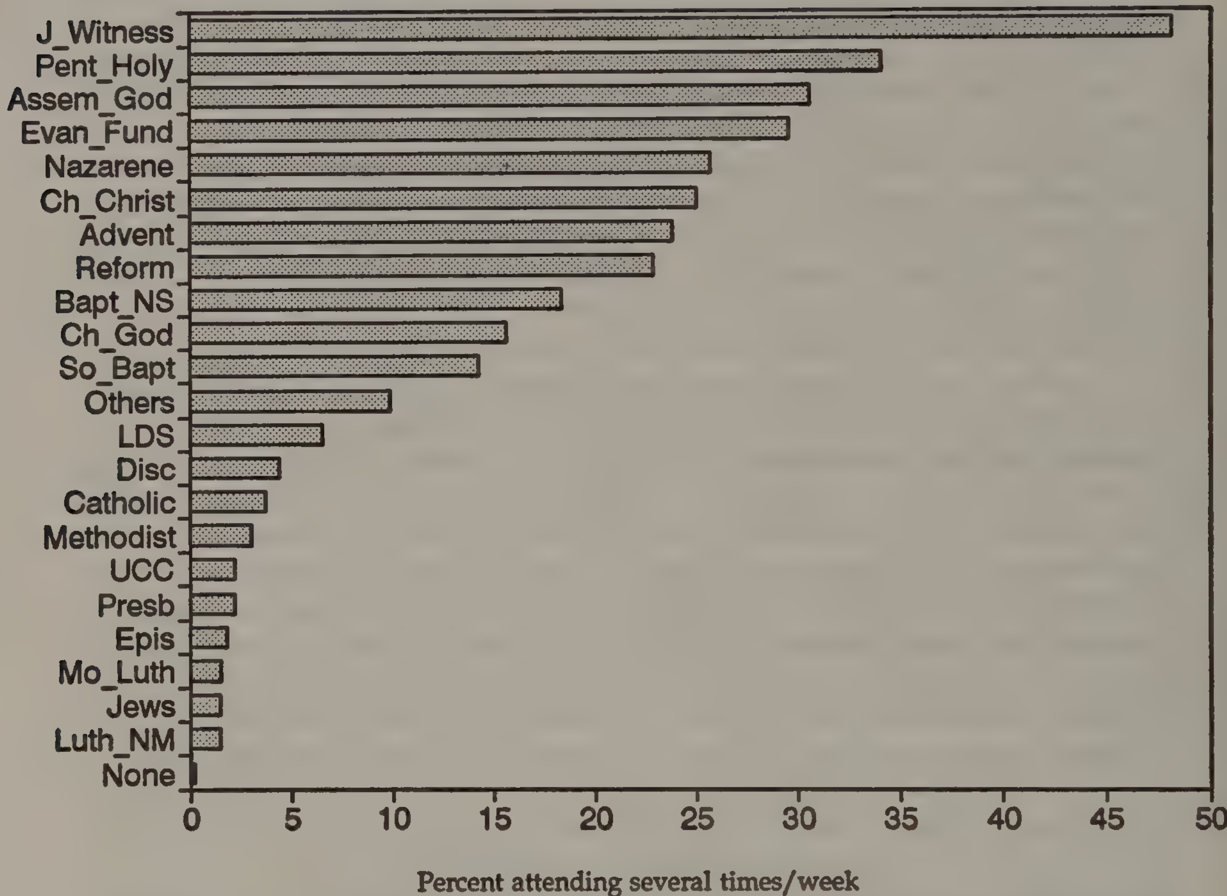
Although more detailed data certainly would help, these figures leave no doubt that many of Kelley's "strict churches" have at their disposal vastly more volunteer labor than their mainline counterparts, per congregation and per capita. For all intents, there simply is no one minding the mainline store six days a week.

It is possible, of course, that the mainline makes up in dollars what it lacks in volunteer hours. Indeed, the proposed model of growth explicitly acknowledges substitution between labor and capital or time and money. Lay denominations, like the Mormons and the Jehovah's Witnesses, require substantial amounts of volunteered time to compensate for their lack of a paid clergy. Mainline denominations depend heavily upon paid and professional services, not merely in the pulpit but also in their administration, counseling, missionary societies, educational programs, music ministries, and building maintenance. Mainline members also average higher incomes and wage rates than their sectarian counterparts, so they have a rational incentive to pay for services (through contributions) that poorer people might produce with their own time.

Until quite recently, cross-denominational survey data on religious giving was largely unavailable.[8] However, in 1987 the General Social Survey began asking respondents "About how much do you contribute to your religion every year (not including school tuition)?" Though this question is not without problems (most notably, in failing to distinguish individual and household contributions), it provides considerable insight into denominational differences. Figure 2 graphs the average number of dollars contributed per (self-identified) member across various denominations. Once again, the mainline denominations place well below their sectarian counterparts. The contrast becomes even more striking when we measure *percentages* of income contributed to one's church. The Episcopalians, Disciples, Lutherans, Methodists, Presbyterians, UCC members, and Catholics all contribute less than 2% of their income. Southern Baptists, Adventists, fundamentalist/evangelicals, and pentecostal/charismatics contribute between 2% and 4%. And Mormons contribute more than 5%, nearly three times the mainline average.

Expert judgments provide a final approach to assessing resources. As an alternative to the survey-based measures of time and money, we asked sixteen scholars doing research on American religion to rate denominations according to their "demands on members' resources."[9] Specifically, we asked each scholar to consider: "Does the denomination expect members to devote a great deal of their time and/or money resources to the denomination's activities or causes?" They then assigned each of twenty-one Christian denominations a score from one to seven, with one representing the lightest demands and seven the heaviest. The level of agreement among the experts proved extraordinarily high. Interrater reliability of denominational scores (as measured by Cronbach's alpha[10]) was .98. The ratings also correlate strongly with the previously

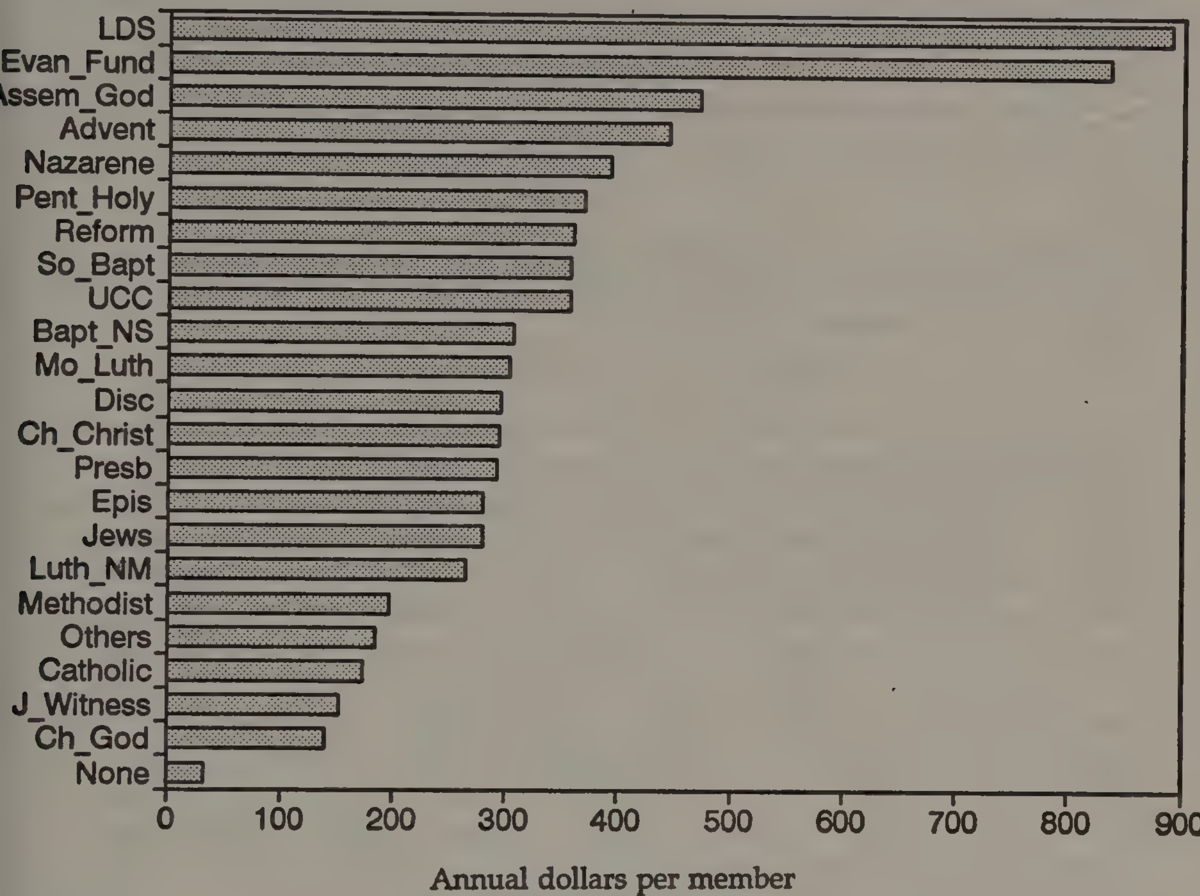
FIGURE 1: Weekday Church Attendance[a]



[a] Source: General Social Surveys, 1984-1990. See Roof and McKinney (1987:253-56) for details concerning denominational classifications.

Advent	=	Adventist
Assem_God	=	Assemblies of God
Bapt_NS	=	Baptist (not Southern)
Catholic	=	Catholic
Ch_Christ	=	Churches of Christ
Ch_God	=	Church of God
Disc	=	Christian Church (Disciples of Christ)
Epis	=	Episcopalian
Evan_Fund	=	Various Evangelical/Fundamentalist groups
J_Witness	=	Jehovah's Witnesses
Jews	=	Jewish
LDS	=	LDS and RLDS Mormon
Luth_NM	=	Lutheran (not Missouri Synod)
Methodist	=	Methodist
Mo_Luth	=	Missouri Synod Lutheran
Nazarene	=	Nazarene
None	=	No religion
Others	=	All other Christians
Pent_Holy	=	Various Pentecostal groups
Presb	=	Presbyterian
Reform	=	Reformed Christian Churches
So_Bapt	=	Southern Baptist
UCC	=	United Church of Christ

FIGURE 2: Church Contributions[a]



[a] See Figure 1 for key of abbreviations.

described measures of weekday attendance and contributions (yielding a multiple correlation of .85 and simple correlations of .65 and .51, respectively). We return to these ratings in the next section.

Predicting Growth

Having used survey data to highlight the differences between American denominations in the resources available to them, we now use these same data to predict denominational growth. Stated differently, we test the time-and-money model of church growth using the only representative, cross-denominational sample currently available. If this simple model and crude explanatory variables account for a significant fraction of the observed variation in denominational growth, then one can be fairly confident that the resource-oriented approach merits further attention.

Predicting denominational growth with national survey data means working with sixteen cases in the 1970s and fifteen cases in the 1980s. The smaller denominations simply do not have adequate numbers of respondents in the GSS even after merging the survey responses from several years. Degrees of freedom thus limit the analysis to simple models with one or two independent variables.

Linear production functions provide the simplest approach to modeling growth as a function of time and money inputs:

$$G = a + b_l * L + b_k * K$$

Alternatively, one may employ an equally parsimonious loglinear specification (known as the "Cobb-Douglas" function) popular in economics and used extensively by Coleman (1990:667-949):

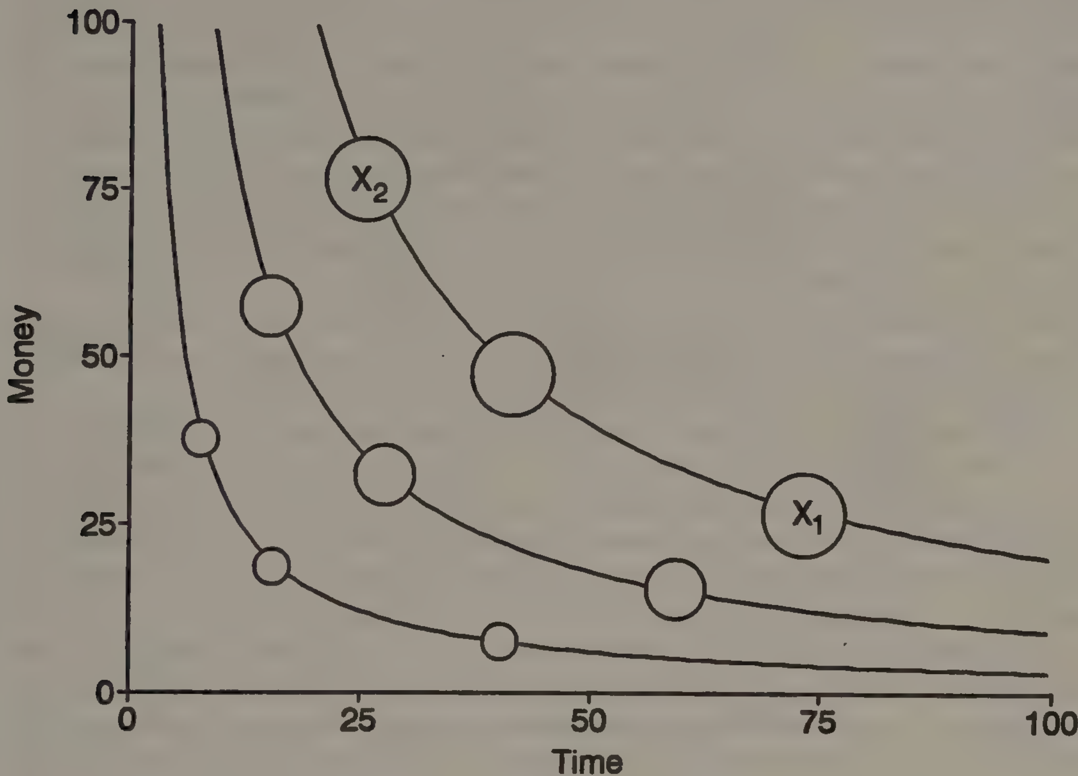
$$\log G = \alpha + \beta_l * \log L + \beta_k * \log K$$

Dividing all three variables by total membership, one obtains equations that describe growth *rates* as functions of time and money *per capita.* Ordinary least-squares regression can be used to estimate either set of equations. Although the linear specification appears somewhat simpler, resource mobilization scholars might nevertheless prefer the Cobb-Douglas specification, because it limits substitutability among inputs (and thereby maintains a substantive distinction between time and money). The Cobb-Douglas equation implies that each 1 *percent* increase in L leads to a β_L% increase in G, and each 1% increase in K leads to a β_K% increase in G. Hence, successive reductions in time can be offset only by progressively larger amounts of money, and no amount of money can take the place of all the time. In contrast, the additive structure of the linear model implies that each 1 *unit* increase in L inputs yields a b_L-unit increase in G. Time and money thus act as perfect substitutes, and each one-unit reduction in L can always be offset by exactly b_K/b_L more units of K.

Production functions can be depicted graphically using "isoquant" diagrams. Figure 3 displays an isoquant diagram corresponding to a Cobb-Douglas function. For each point in this diagram, such as the point labeled x_1, the horizontal and vertical coordinates (75,25) show the amount of time and money dedicated to production and the size of the circle surrounding the point depicts the resulting level of output. Since every combination of time and money yields *some* level of output, one may associate an output circle with every point in the diagram. However, the general shape of the production function may be inferred from just a few such points corresponding to a few levels of output.

In Figure 3 the innermost set of circles (closest to the origin of the time and money axes and thus using low inputs of time and money) depict input combinations that all yield the same, relatively low, output. The outermost set of circles depict input combinations that yield a high output. The intermediate circles denote input combinations that yield an intermediate level of output. The entire set of points that yield a particular level of output is called an isoquant. From the shape and position of the isoquants, one can infer the character of the underlying production process. In particular, one can draw inferences about *substitution* between inputs. In Figure 3, the points X_1 and X_2, depict two radically different methods of generating the same result, X_1 being much more *labor-intensive,* since it employs a much higher ratio of time to money.

FIGURE 3: Growth Isoquants



Turning to data, the *Yearbook of the American and Canadian Churches* (Jacquet & Jones 1992) provides annual reports of denominational membership (the membership figures having been supplied by the denominations themselves). The variable *growth 1980-90* measures a denomination's membership in 1990 divided by its membership in 1980 and so gives the proportional change over the last decade. The preceding section's discussion of religious resources explains why standard church attendance rates tend to overestimate the labor resources available to a denomination. Hence, we estimate *time* as the fraction of a denomination's members who attend religious services "several times a week." We estimate *money* by averaging the (inflation adjusted) amount of money that the members of each denomination reported in response to the GSS contribution question. Comparable measures may be defined and calculated for 1970 through 1980.[11]

The first two columns of Table 1 fit this simple, two-variable regression quite well, having an adjusted R^2 of .68 in the 1980s and .81 in the 1970s. Both sets of explanatory variables have the expected positive signs, and both are statistically significant. It would seem that religious resources do indeed "produce" membership growth in accordance with standard production models. Although these regressions employ a linear functional form and specific measures of time and money, the results prove quite robust. The logarithmic

counterparts of all these equations yield qualitatively similar results, as do regressions using alternative measures of attendance and contributions.[12]

Figure 4 graphs the relationship between attendance, contributions, and church growth for the twenty denominations. Following the style of Figure 3, the horizontal and vertical axes measure a denomination's time and money inputs and the size of a denomination's circle depicts its growth rate. The small circles clustered in the lower left of the diagram represent the mainline denominations: low inputs and low to negative growth (averaging between zero and -15%). Most of the conservative denominations and sects occupy the middle of the diagram: their relatively high inputs yield moderate to high rates of growth (averaging between 10 and 20%). Two highly sectarian denominations — the Mormons and Jehovah's Witnesses — stand out in their exceptionally high rates of growth (over 50%) and high levels of attendance or contributions. Jehovah's Witnesses distinguish themselves as by far the most labor-intensive denomination; Mormonism is extraordinarily money-intensive.

The third column of Table 1 shows that expert judgments about demands on denominational resource predict 91% of the variation in denominational growth rates in the 1980s and 74% in the 1970s, more than the survey-based time and money measures combined.[13] Figure 5, which graphs the relationship between resource demands and growth in the 1980s, is nothing short of amazing. In contrast, columns 4 through 7 of Table 1 show that standard demographic variables do a much poorer job of predicting growth. Other demographic controls, such as race, sex ratios, and percentage married do worse yet. The small number of observations makes it impossible to estimate resource regressions while simultaneously controlling for demographic variables. But in stepwise regressions, resource demands and time variables consistently entered before demographic variables, and their *t*-statistics did not decline appreciably after controlling for individual demographic variables.[14]

The basic message is clear. In each of the past two decades, the decline of the liberal/mainline denominations and the growth of their more strict/conservative counterparts can be modeled largely in terms of the radically different amounts of time and money that they demand and receive from members. At this level of analysis, denominational dynamics prove surprisingly straightforward.

From Denominations to Congregations

Although the discussion so far has centered on denominations, there is nothing in the resource model that requires its application to aggregate levels. On the contrary, the model's growth-producing processes tend to operate at the level of individual congregations as sermons are preached, classes are taught, chapels are maintained, and music is performed. Hence resources should also predict the growth of congregations. If the model fails at the congregational level, one might reasonably question whether the denominational findings are artifacts of spurious correlation and the small number of observations.

We have, in fact, found two distinct congregational data sets suitable for analysis. Both permit tests of the effects of the resource variables while also

TABLE 1: Predicting Denominational Growth[a]

	growth 1980-90	growth 1970-80	growth 1980-90	growth 1970-80	growth 1980-90	growth 1970-80
time80	.6867*** (4.55)	...	...	...	...	...
money80	.5525** (3.66)	...	...	...	...	...
time70	...	.6591*** (4.42)	...	...	...	...
money70	...	.3349* (2.25)	...	...	...	...
demands	...	...	.9584*** (12.11)	.8729*** (6.70)	...	...
income80	...	...	...	...	.6867 (0.40)	...
educ80	...	...	...	...	-.0771 (0.18)	...
age80	...	...	...	...	-.8453** (2.60)	...
income70	...	...	...	...	...	-.3790 (0.82)
educ70	...	...	...	...	...	.3349 (0.10)
nchild70	...	...	...	...	...	.4144 (1.83)
Cases	15	16	15	16	15	16
Adj. R^2	.683	.806	.912	.740	.421	.350

[a] Sources: General Social Surveys, 1972-1990; Yearbook of American and Canadian Churches, 1970-1992; and a survey of experts.

Beta values are in plane text; absolute *t* statistics are in parentheses.

Variable definitions: All for the following variables are measured or aggregated at the level of the denomination.

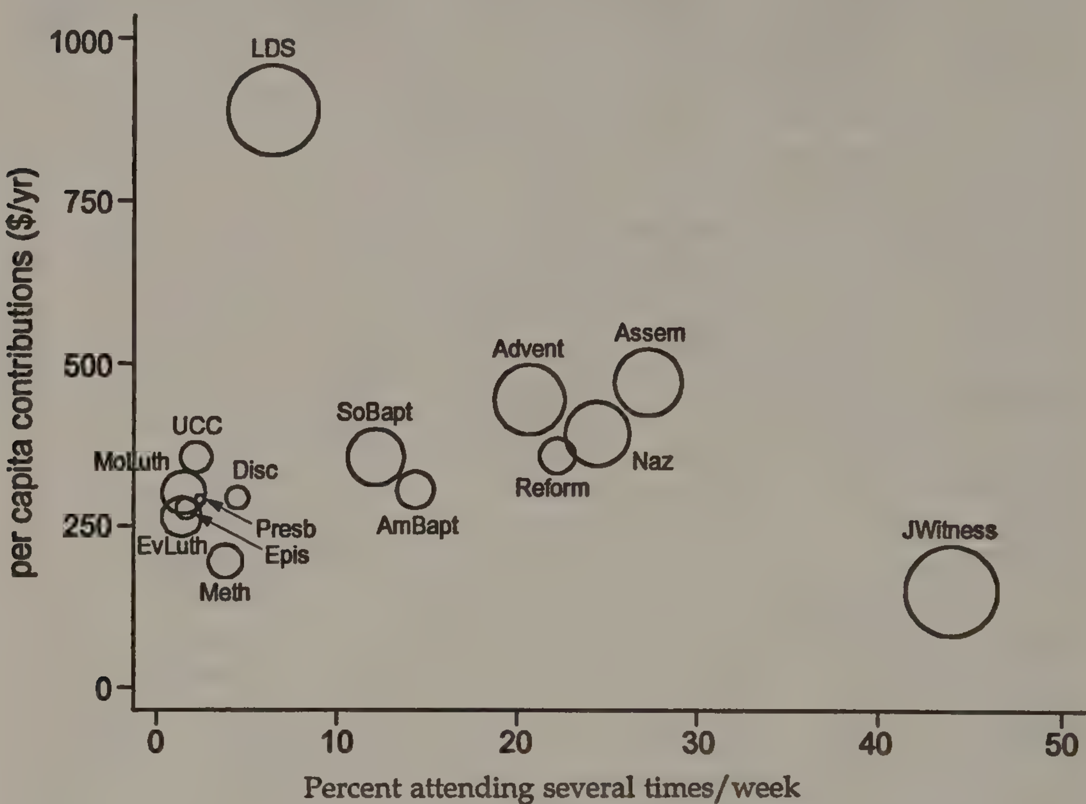
Yearbook data: growth 1980-90 = 1990 membership divided by 1980 membership; growth 1970-80 = 1980 membership divided by 1970 membership; money70 = contributions per member.

Expert ratings: demands = 7-point, additive scale.

GSS data: time80, time70 = percent attending "several times per week"; money80 = (inflation adjusted) average contribution, 1987-1989; income80, income70 = (inflation adjusted) real income; educ80, educ70 = average years of education; age80 = average age; nchild70 = average number of children.

* $p < .05$ ** $p < .01$ *** $p < .001$

FIGURE 4: Denominational Growth, 1980-1990

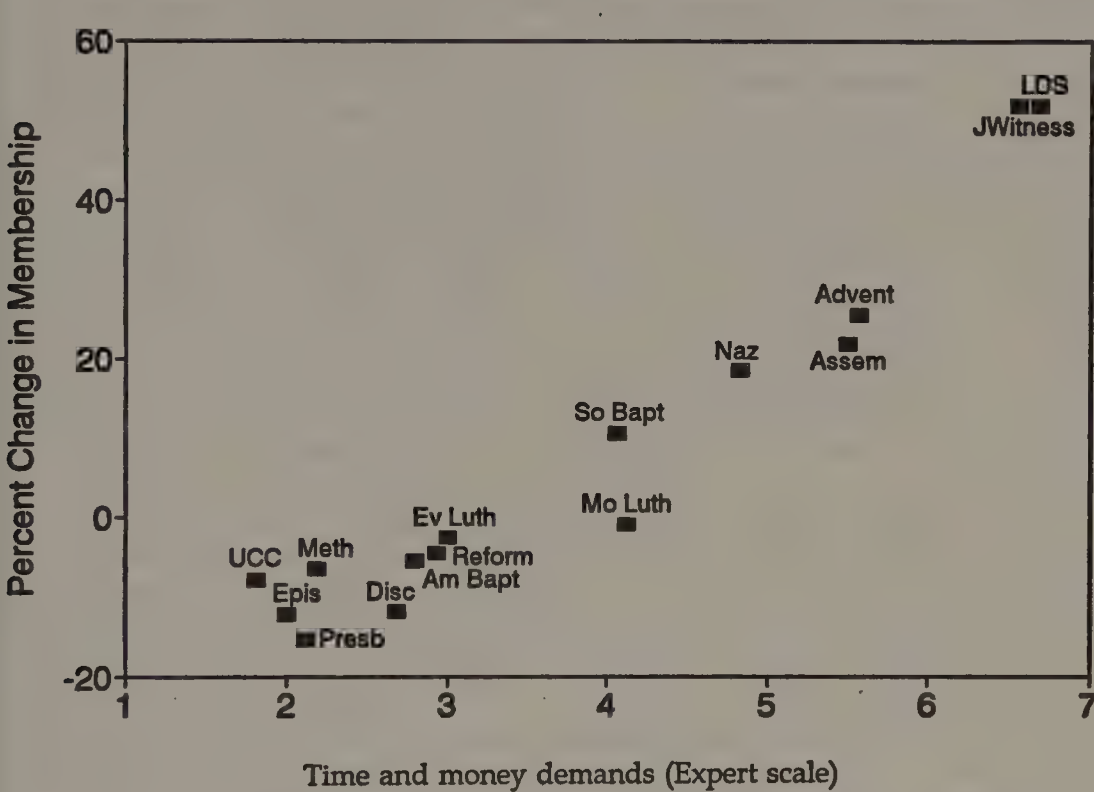


controlling for the effects of variables other than resources, and both confirm the importance of resources.

The first set, which covers about 400 Disciples of Christ congregations, was assembled and analyzed by Meyers and Olson (1991). The data include congregational membership from 1979 to 1983 and a variety of other congregational attributes measured in 1979. Using regression techniques and the variables at hand, Meyers and Olson sought to identify the statistical predictors of church membership growth. They concluded that "high levels of member involvement (as reflected in higher levels of per capita giving and higher rates of participation in church school) *precede* increases in church size" (509). To state the matter bluntly: time and money predict growth better than virtually any other variables.

Table 2 reprints Meyers and Olson's key regression. A few points deserve emphasis. First, the dependent variable is the rate of change in a congregation's membership from 1978 through 1983. Second, the full set of independent variables is quite diverse, and all of them predate the period over which growth is measured. Third, two of the three strongest predictors (as measured by t statistics and betas) turn out to be expenditures per member and frequency of church school attendance per member. The expenditure variable constitutes the only available measure of money inputs, since individual giving was not part of the data set. Church school attendance constitutes the only available proxy for time inputs. It gives the average number of children and adults attending

FIGURE 5: 1980-1990 Membership Growth versus Resource Demands



Sunday school and Bible studies and so provides a measure of participation beyond the Sunday worship service. As can be seen in Table 2, several other congregational attributes prove statistically significant, but none except race is as powerful as these two resource-oriented measures. For further details, see Meyers and Olson (1991).

Our second set of results come from a much larger study conducted by the United Church of Christ, a liberal mainline denomination that has been losing members since the 1960s. The data cover nearly a thousand UCC churches and include both congregational and individual responses. The individual-level data were collected as part of a 1975-78 "Church Membership Inventory" in which each participating church distributed a 43-item questionnaire to all persons in attendance during a regular worship service. According to the project director, William McKinney, virtually everyone present completed the questionnaire, yielding a total of more than 70,000 responses and an average of 77 respondents per church. We used these individual responses to compute aggregate values (such as the respondents' average age and income) for each church. We also merged in data from the UCC's 1973-81 yearbooks, which include official membership and financial statistics for each church. Finally, we obtained decennial census statistics for the communities surrounding each church.

As in the denominational regressions, growth is defined as the number of members in one period (1981) divided by the number of members in an earlier period (1977). Since nearly all the participating churches completed their

TABLE 2: Predicting Congregational Growth Disciples of Christ, 1979-1983

	growth
sunday_school	.190 (3.34)***
expense	.293 (5.08)***
elite	.106 (1.87)
race	-.238 (4.01)***
non_disciple	-.120 (2.11)*
non_seminary	-.124 (2.09)*
cases	272
Adjusted R^2	.185

[a] Source: Meyers and Olson (1991: 513).
Beta values are in plane text; absolute *t* statistics are in parentheses.
Variable definitions:
growth = percentage growth in congregation, 1979-1983;
sunday_school = proportion of members who were in church school in 1979;
expense = per capita congregational expenses in 1979;
race = indicator variable for predominantly white membership;
elite = proximity to a Disciples college or seminary;
non_disciple = the proportion of the last 12 years during which the congregation was led by person(s) with a degree from a mainline, disciple-related, but not disciple-owned seminary;
non_seminary = proportion of years led by person(s) with no seminary training.

* $p < .05$ ** $p < .01$ *** $p < .001$

questionnaires before the end of 1977, this measure of growth postdates the other congregational measures, and thus cannot have influenced the independent variables of our regressions.

Unfortunately, membership data for congregations are notoriously "noisy" because the rolls are cleaned of inactive members only irregularly.[15] This lowers the R^2 values obtained in regressions, but it also means that only the strongest independent variables appear as significant predictors of congregational growth. If resource measures predict growth in even these noisy data, one can be fairly confident that they really do matter.

From the UCC yearbook statistics we have computed two measures of monetary resources: per capita contributions and per capita church expenses. (The yearbook statistics exclude bequests, endowments, and income from other sources, such as building rental and product sales.) These data reflect the actual money received by the church, and so avoid biases due to mistaken or exaggerated reporting.

We have already noted that church attendance provides only a crude measure of the time that members devote to their churches. Sadly, the UCC data do not provide a very good measure even of attendance. The average attendance reported by questionnaire respondents contains a strong upward bias, in that regular church attenders were much more likely to be in church when the questionnaire was distributed. Hence, we have estimated attendance as the number of questionnaire respondents divided by the same year's number of official members.[16] The number of questionnaires provides a reasonable estimate of attendance on a typical Sunday, since they were distributed to and completed by nearly everyone in attendance.

The first column of Table 3 contains the basic, linear regression results. As in the denominational data set, contributions and attendance significantly predict growth. UCC congregations with higher per capita contributions and higher attendance (relative to membership) manifest significantly more membership growth in the four years following the survey. In other words, contributions and attendance act as leading indicators of congregational growth. The regression's adjusted R^2 is low, around 6%, though it is hard to imagine a good fit emerging from such noisy measures of growth and such crude measures of participation.[17] There are, moreover, numerous environmental factors that can influence growth at the local level.

The second and third regressions, in columns 2 and 3 of Table 3, give these other factors a chance to compete with the resource variables. Though cognizant of the dangers of statistical fishing expeditions (Leamer 1978), we nevertheless used one here simply to show that the estimated impact of resources is not an artifact of omitted variables. Column two reports the results obtained in a stepwise search over dozens of nonresource variables from the questionnaire, census, and yearbook data. All variables are measured or averaged at the level of the congregation.

Only three nonresource variables proved strongly related to congregational growth. As we have seen in previous research (e.g., Hoge & Roozen 1979a), changes in the surrounding community's population significantly affect congregational membership. Our results also echoed previous research in that population transiency (proxied in our data by the proportion of questionnaire respondents renting an apartment) proved a strong negative predictor of future growth. In the language of the production model, transiency probably causes the membership to "depreciate" faster. Following Olson (1987, 1989), we find evidence that a very high level of average membership tenure also inhibits growth. In churches filled with long-time attenders, current members have many friends, but newcomers find it difficult to form fellowship ties and so perceive the churches as "closed" and "cliquish." In the logarithmic, but not the linear, versions of the regressions, membership growth before 1977 correlates negatively with subsequent growth (suggesting that growth in one period tends

TABLE 3: Predicting Congregational Growth United Church of Christ, 1977-1981

	growth	growth	growth
contributions	.154*** (4.45)	...	.138*** (3.86)
attendance	.157*** (4.54)	...	.110*** (2.92)
renters	...	-.130*** (3.75)	-.143*** (4.13)
yrs_attend	...	-.159*** (4.28)	-.102 (2.66)
pop_growth	...	.148*** (4.03)	.134 (3.70)***
growth7377	...	-.013 (0.37)	.043 (1.24)
members77		-.096** (2.91)	-.031 (0.85)
Cases	842	842	842
Adjusted R^2	.058	.095	.123

[a] Beta values are in plain text; absolute t statistics are in parentheses.
Variable definitions: growth = 1981 membership/1977 membership;
contributions = per capita contributions;
attendance = number of questionnaire respondents/1977 membership;
renters = share of members living in rented housing;
yrs_attend = average years attending this congregation;
age = average age of respondents;
growth7377 = membership in 1977/membership in 1973;
members77 = congregation membership in 1977
pop_growth = community population in 1980/community population in 1970.
All variables are measured or aggregated at the level of the individual congregation.

* $p < .05$ ** $p < .01$ *** $p < .001$

to dissipate the resources needed to facilitate future growth — see Meyers & Olson 1991). Finally, larger congregations tend to have lower growth rates, but this effect turns insignificant after taking account of contributions and attendance (suggesting that insofar as initial size may affect subsequent growth, it does so through its impact on participation levels). Altogether, the nonresource measures account for about 9.5% of the variance in growth. Variables dropped for lack of statistical significance were the demographic characteristics of

individual church attenders (average age, average income, average education, average occupational prestige, average number of children living at home, etc.), homogeneity-diversity of church attenders (standard deviation of ages, incomes, education, etc.), percentage distribution of the church's overall budget to different categories (capital expenditures, money sent to the denomination, and money sent to nondenominational organizations like World Vision or local charities), and attenders' average satisfaction with nine aspects of church life.[18]

Comparing columns 1 and 2, we see that the two *preselected* resource variables predict nearly as well as all other statistically significant variables (searched from among several dozen variables) combined.[19] Comparing columns 1 and 3, we see that the statistical influence of contributions and attendance remains strong even after including the other variables. And columns 2 and 3 show that resources increase total explained variance even after including all other significant predictors. In unconstrained stepwise regressions, contributions and attendance entered before any variable except population growth. The logarithmic counterparts of all three regressions yield qualitatively similar results.[20]

In short, resources really do matter. Whether we study entire denominations or individual congregations, time and money stand out as statistically significant determinants of growth. The theoretical model survives empirical scrutiny — a result with general implications for organizational analysis and specific applications to religious research.

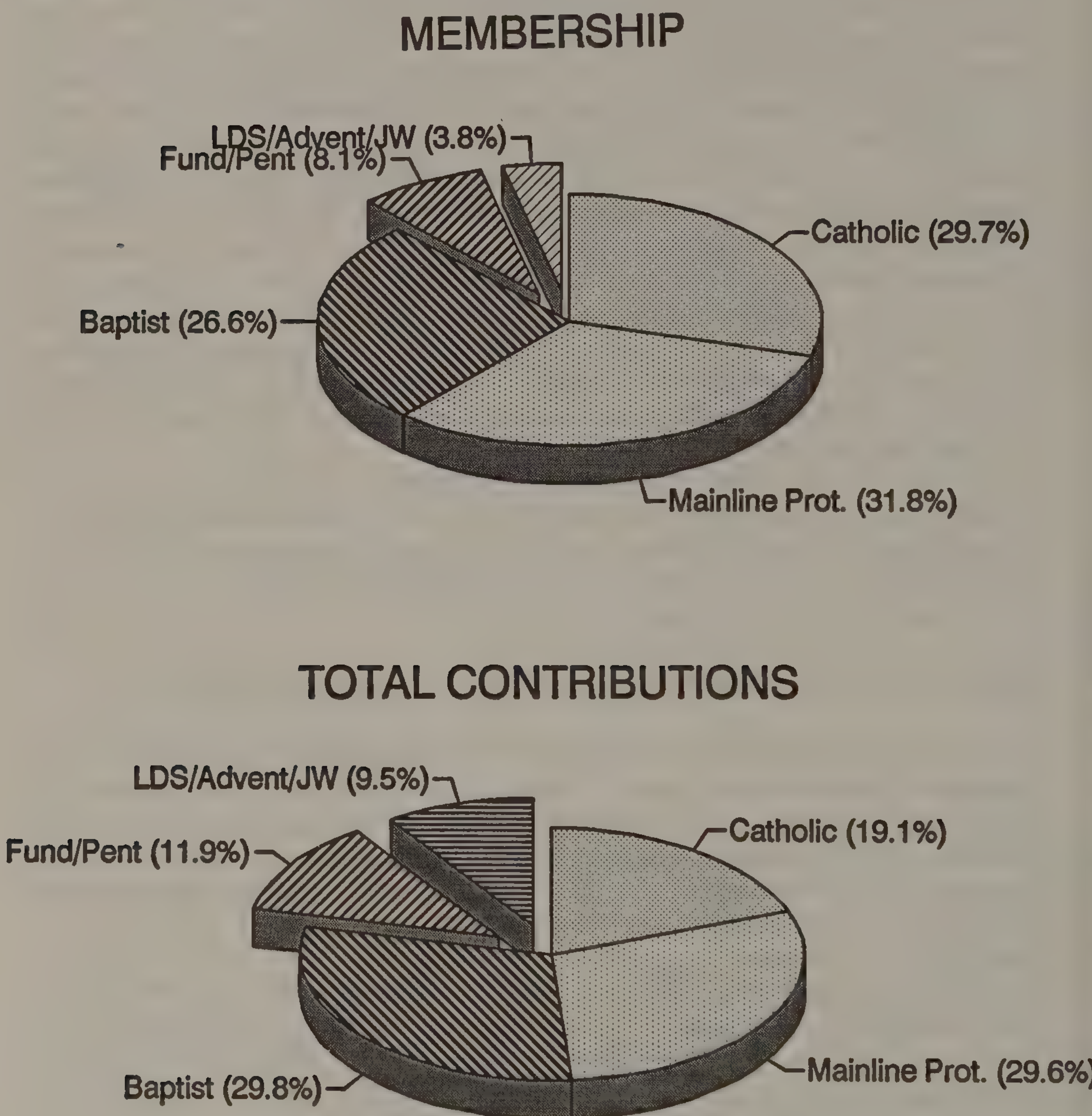
Extensions: "Resourceful" Thinking about Religious Market Shares

Having demonstrated that resources affect church growth at both national and local levels, we turn briefly to a related issue — assessing the relative sizes of denominations. Our goal is twofold: to further demonstrate the value of resource-oriented thinking and to provide some specific insights that call for a major reevaluation of the relative size and strength of American denominations and a reassessment of some longstanding notions about the difference between Catholics and Protestants.

One hears much talk these days about "religious markets." Indeed, Steven Warner (1993) recently argued that the market-oriented approach has attained the status of a "new paradigm" in the sociology of religion and is well on its way toward displacing the traditional secularization paradigm. Even so, scholars have not given much thought to measuring the size and composition of religious markets.

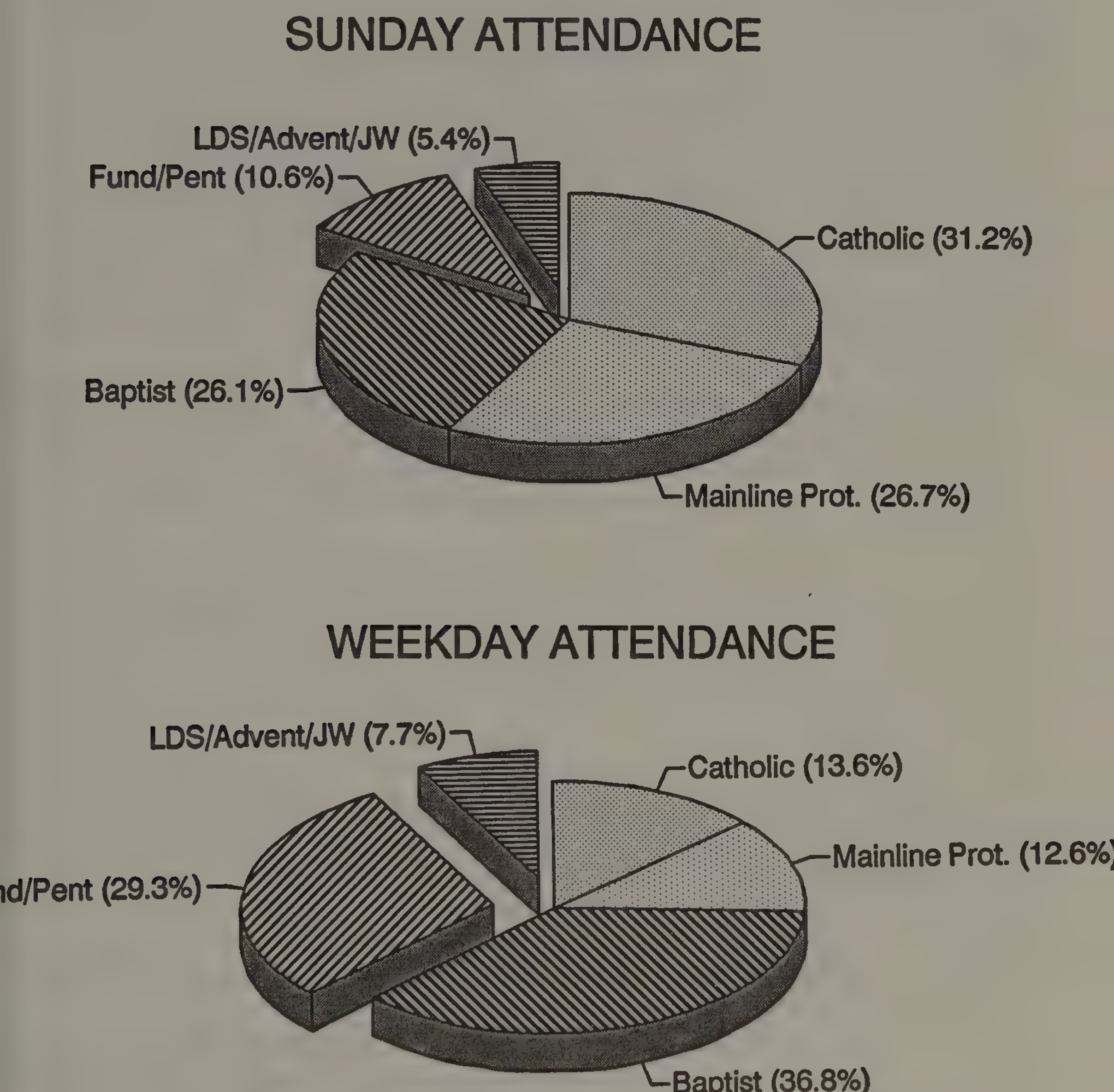
In standard markets, such as the market for automobiles, industry analysts routinely calculate market shares in both dollar and "unit" sales. Religious research would benefit greatly from an analogous distinction. Traditional scholarship has, more or less unconsciously, based its conclusions on *numbers* of official or self-identified members. Membership numbers have their uses, but they often prove misleading when comparing groups whose members manifest significantly different rates of participation and commitment. Dollar contributions provide an alternative measure of market share, and probably a more useful measure of market power. The total amount of time devoted to religious activities provides yet another measure.

FIGURE 6: Alternative Measures of Market Share[a]



The relative market shares of major denominations change dramatically with the measure we employ. For example, the Methodists and Southern Baptists have fairly similar numbers of self-identified members, but the contributions of the latter are nearly *twice* those of the former. Measured in dollar contributions, the Mormon church is virtually the same size as the Methodists, despite a total membership that is one-seventh their size. The Jehovah's Witnesses have more members attending "several times per week" than are found among the Episcopalians, UCC, and Presbyterians combined! (The mainline denominations' share of the financial pie shrinks further still if

FIGURE 6: Alternative Measures of Market Share[a] (Continued)



[a] Source: General Social Surveys, 1984-90
Fund/Pent = Fundamentalist and Pentecostal groups
LDS/Advent/JW = Mormons, Adventists, and Jehovah's Witnesses

one subtracts monies spent on clergy and staff before making comparisons with groups like the Jehovah's Witnesses and Mormons whose lay clergy guarantees that most contributions go to missionary work, building construction and maintenance, and other church programs.)

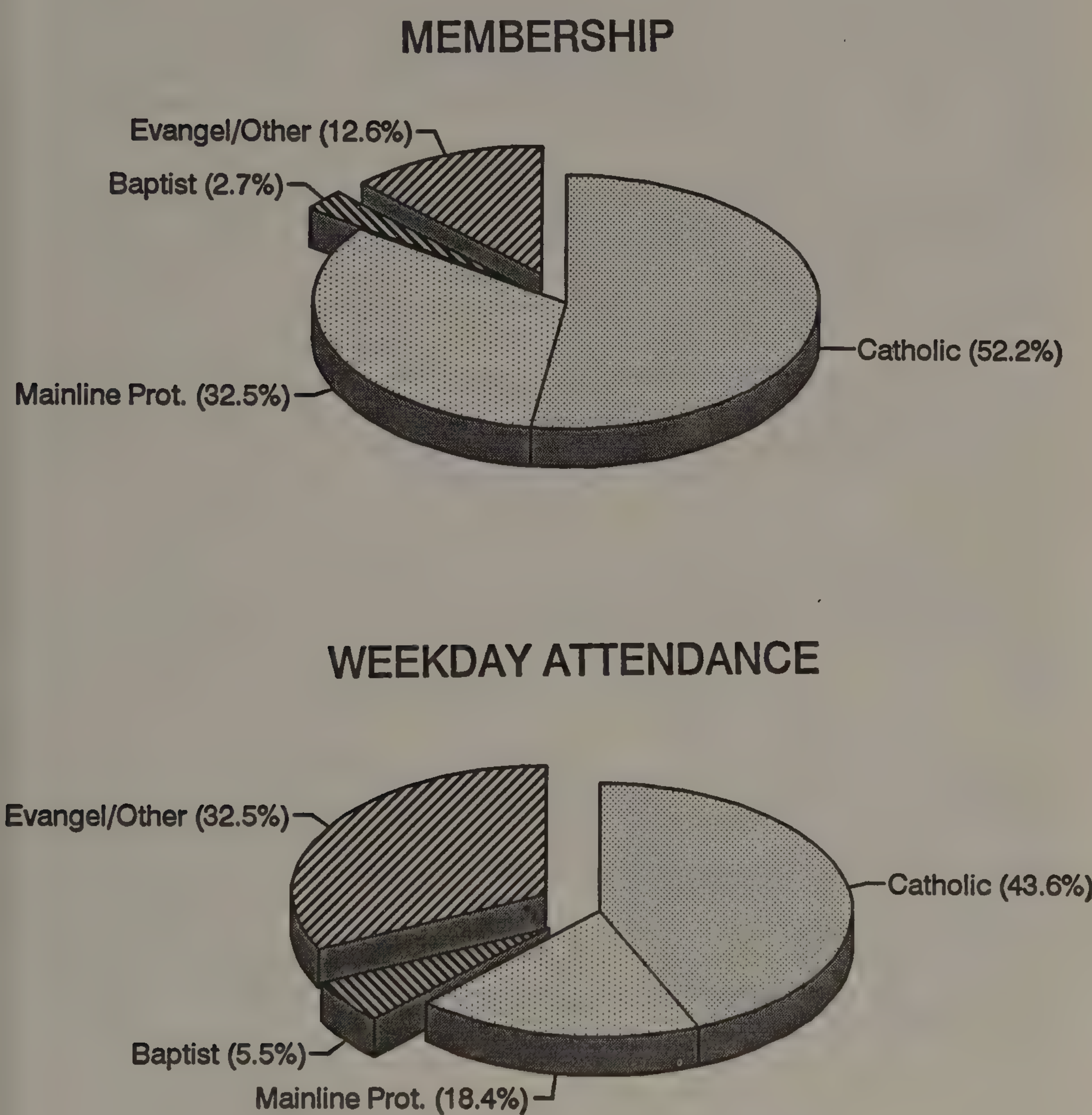
For an overview of these differences, consider the four pie charts in Figure 6. The charts provide four views of the market shares of Christian denominations — membership, Sunday attendance, weekday attendance, and total contributions. All data come from the 1984-90 *General Social Surveys,* and the only omitted groups are "nondenominational" Protestants and a handful of very small, hard-to-classify denominations. In the market for (self-identified) members, Catholics, non-Baptist mainline Protestants, and Baptists each have approximately 30% shares, whereas the sects (Fundamentalists, Pentecostals, Mormons, and others) have a mere 12%. Measured in terms of Sunday church attendance, the mainline shrinks a bit and the sects rise to 16%. Measured in terms of dollar contributions, the sects rise to more than 20%, and Catholics shrink to less than 20%. In other words, despite having three times the population, Catholicism's share of the market for religious contributions is no larger than that of America's sectarian "fringe." The sectarian share is all the more striking given that their members average relatively low incomes.

The most dramatic result concerns weekday attendance, defined here as the number of GSS respondents claiming to attend religious services "several times a week." By this measure, sectarians constitute more than a third of the market, as much as the Baptists, and substantially more than the Catholics and the mainline Protestants put together.

Data from Canada's *General Social Survey* prove that the preceding results are not unique to the U.S. The 1986 Canadian GSS asked people whether they had attended church any time Monday through Friday of the past week. Figure 7 shows how the response to this question differs from the response to the more standard membership question. Although the Canadian survey's denominational breakdown is less detailed than that obtained from the U.S. data, the basic pattern is the same. Conservative Protestants (Baptists and Evangelicals) make up only 12.8% of the market for members — less than half the size the more liberal Protestant denominations (Anglican, Lutheran, UCC, and Presbyterian) and less than one third the size of the Roman Catholic membership. But their share of the weekday market is 38%, just shy of the Catholic share and more than twice the liberal Protestants' share. Hence, in Canada as in the U.S., simple measures of membership or affiliation provide a biased view of the religious marketplace.

Why focus on the relatively small number of people who attend church "Monday through Friday" or "several times per week"? Because it is precisely this core of very committed members who provide the surplus labor that makes church growth possible. In like manner, future research should focus attention on the surplus dollars available for expansion, evangelism, and renewal, a surplus left after subtracting the portion of contributions that merely cover basic operating costs. The preliminary evidence (of Figures 2 and 6) leave little doubt that sects command a disproportionate share also of this pie.

FIGURE 7: Market Shares in Canada[a]



[a] Sources: Canadian General Social Survey, 1986
Evangel/Other = Evangelicals and other Christians

Conclusions

Why do some religious organizations grow while others decline? The complete answer is, no doubt, exceedingly complex. Nevertheless, both theory and data suggest a strikingly simple partial answer. Growth requires resources. It is a "product" derived from "inputs" of labor and capital, which is to say time and money. Hence, religious organizations with high average rates of participation tend to grow, and religious organizations with low average rates of participation tend to decline.

Empirical research on church growth has often emphasized every factor *except* attendance and contributions — theology, polity, demographics, attitudes, population growth, and the state of the economy. Our analysis suggests that the impact of these variables is dwarfed by that of time and money. A shift in scholarly focus would seem to be in order. Resources deserve more attention and much more careful measurement. Previous research has tended to view church attendance as a simple attribute, comparable to an individual's beliefs or demographic characteristics. We have argued that church attendance is merely the tip of the resource iceberg. Regular attendance is well and fine, but not to be confused with the surplus resources that an organization needs to thrive. Researchers need to look closer, so that it will be possible to speak not only of the amount of time that members devote to an organization but also the particular activities they have undertaken. Previous research on contributions is even more limited than the work on time; indeed it is almost nonexistent. Our results underscore the need to devote much more attention to the measurement and analysis of religious giving.

In the end, however, resource-oriented research complements, rather than replaces, more traditional approaches. It especially complements existing research on the determinants of denominational strength and success. The theoretical link from organizational strictness to organizational strength has received a great deal of attention in recent years. The empirical relationship between strictness and growth has been observed for more than two decades, but causal explanations have been largely lacking. Resources provide a critical bridge, helping us see how a strict organization's strength leads (via high rates of participation) to growth.

Notes

1. Many "New Religious Movements" are exceptions to this rule. Bromley (1985), Lofland (1979), and others have effectively employed standard resource mobilization theory to study the financial strategies of the Unification Church and other NRMs (see Robbins 1988:127-33).

2. When volunteers provide "free" services, they bear an opportunity cost equal to the value of their sacrificed time and energy. The church also bears a cost, since the person who volunteers to do one job cannot simultaneously do other church jobs.

3. Strictly speaking, a standard economic growth model would look more like the following:

$$dQ/dt = F(L_t, K_t) - \delta \cdot Q_t$$

where Q_t represents quantity of members at time t, δ represents the rate at which membership "depreciates" or declines (due to deaths, dropouts, mobility, depreciation of facilities, and the like) in the absence of continuing investment, the derivative dQ/dt represents the

current rate of membership growth, and L_t and K_t represent the flows of labor and capital invested in continued growth.

4. It is, of course, possible to collect census, survey, and church record data on hundreds of factors that might fall into one or more of the preceding headings, and then search through massive correlation matrices and stepwise regressions for items that best predict growth. Most empirical studies of congregational growth do just that (Hoge and Roozen 1979a; Roozen and Hadaway 1993). Such methods have value in exploratory research, but they do not constitute hypothesis testing, and they do not take the place of formal models. For more on this subject, see Iannaccone (1996).

5. The model does not rule out the impact of the other factors noted above; it merely assumes that they operate indirectly. Theology, socioeconomic attributes, commitment, goals, and the social environment can still influence membership retention and recruitment, but they must do so *through* resources. For example, a member's personal feelings of commitment can affect growth, but only insofar as these feelings give rise to concrete actions — reaching out to people in need, serving on church committees, testifying publicly to one's faith, and contributing to the building fund. The model thus restricts the mechanism by which other factors determine growth.

6. The percentages in Figure 1 are undoubtedly inflated, since survey respondents tend to overstate their actual rates of church attendance (Hadaway, Marler & Chaves 1993). However, given the widespread nature of this bias, the denominational comparisons almost certainly remain valid.

7. The Mormon rate of 6.4% is misleadingly low because beginning around 1979 the LDS Church consolidated its meeting schedule, moving several types of services, which previously met throughout the week, to Sundays. Hence, although many Mormons now "attend religious services" only once each week, this attendance includes three different meetings totaling about three hours. Moreover, Mormons routinely devote substantial additional time to their religion: visiting other members ("home teaching"), serving on missions, worshiping in the home ("family home evening"), attending a variety of church-sponsored events ("Firesides," dances, and recreational events), and serving the church in administrative, teaching, and other capacities. None of this activity is measured by standard church attendance survey questions.

8. The major exceptions are Stark and Glock's (1968) survey of San Francisco Bay Area church members and a national population survey conducted by the Yankelovich organization for the Independent Sector in 1984.

9. This was only one of nine dimensions along which the experts rated the denominations. The other eight dimensions repeated questions from a study conducted by Dean Hoge (Hoge & Roozen 1979a:179-97, 1979b:E1-E14). The other questions concern the denominations' "emphasis on distinctive life style and morality," "attitude toward pluralism of beliefs among members," "theological conservatism or liberalism," "strength of ethnic identity," "attitudes toward ecumenism," "centralized or congregational polity," "emphasis on local and community evangelism," and "involvement in social action." In all cases, our experts' average ratings correlated almost perfectly ($r > .92$) with those produced by Hoge's experts, suggesting extraordinary stability in the (perceived) character of denominations over the past fifteen years.

10. Typical survey-based scales are formed by summing an individual respondent's scores on several survey items. Cronbach's alpha then provides a measure of inter-item correlation across the sample of respondents. The present scale is formed by summing the responses of 16 experts. Hence, each expert acts as a different "item" or measure of the underlying characteristic (resource demands), and each denomination acts as a separate case. In this context, Cronbach's alpha provides an index of correlation among the experts and thus is a measure of interrater reliability. The presumed statistical model is $s_{ij} = t_j + e_{ij}$, where s_{ij} denotes the ith expert's distinctiveness score for the jth denomination, t_j denotes the jth denomination's true distinctiveness level, and e_{ij} denotes the (random) error in the ith expert's judgment on the jth denomination.

11. The 1970s money data come from denominational figures reported in the *Yearbook*. Ideally, one would like to define and measure all variables on a single, consistent membership criterion. In practice, population surveys assign respondents to denominations from their stated religious preference or identification, whereas denominations report membership figures (and

thus per capita contributions) on the basis of official institutional membership, the criteria for which vary significantly from one denomination to the next. Fortunately, these differences become less problematic when comparing growth rates over time. The effects of "loose" or "strict" membership criteria tend to wash out, because they inflate (or deflate) a denomination's membership figures consistently year after year. Population surveys attain a comparable consistency, since they apply a single, self-identification-based membership criterion across all denominations.

12. The logarithmic regression results corresponding to the second and first rows of Table 1:

$$\log(\text{growth7080}) = \underset{(3.888)^{**}}{.631} \cdot \log(\text{time70}) + \underset{(1.564)}{.304} \cdot \log(\text{money70}) \qquad \text{Adjusted } R^2 = .739$$

$$\log(\text{growth8090}) = \underset{(2.856)^{*}}{.603} \cdot \log(\text{time80}) + \underset{(1.265)}{.267} \cdot \log(\text{money80}) \qquad \text{Adjusted } R^2 = .382$$

13. The experts tended to produce very similar rankings when asked to rate denominations with regard to other attributes — lifestyle distinctiveness, belief strictness, theological conservatism, ecumenism, evangelistic emphasis, and involvement in social action (but not strength of ethnic identity or centralized polity). This might reflect a tendency to oversimplify and implicitly assume that most denominational attributes covary along a single underlying dimension. It also underscores the potential for spurious correlation when working with a small number of observations. On the other hand, there *are* strong theoretical reasons for expecting many of these attributes to cluster (Iannaccone 1994; Kelley [1972] 1986).

14. Simultaneously regressing growth onto resource demands and demographic controls yields:

$$\text{grow7080} = \underset{(4.10)^{***}}{.963} \cdot \text{demands} + \underset{(0.21)}{.066} \cdot \text{income70} - \underset{(0.07)}{.192} \cdot \text{educ70} - \underset{(0.88)}{.184} \cdot \text{nchild} \qquad \text{Adjusted } R^2 = .719$$

$$\text{grow8090} = \underset{(8.04)^{***}}{.985} \cdot \text{demands} - \underset{(1.12)}{.232} \cdot \text{income80} + \underset{(1.75)}{.292} \cdot \text{educ80} + \underset{(0.01)}{.001} \cdot \text{age} \qquad \text{Adjusted } R^2 = .915$$

The impact of demographic controls is equally small when regressing growth onto attendance and contributions. As noted in the text, other controls have even less impact. This includes controls for each denomination's total size and average congregational size. It also includes controls for the regional distribution of the denomination's membership (percentage located in the West, South, and East) and a "regional growth index" (formed as a weighted average of regional population growth rates multiplied by the fraction of a denomination's membership located in that region). For more on the definition of these controls, see Hoge & Roozen (1979b:E1-14).

15. Most UCC conferences set a minimum contribution that each congregation is supposed to send to the denomination. This amount is based on the number of members reported to the denomination, which encourages congregations to remove inactive members. Even so, actual year-to-year changes in reported membership suggest that roll cleaning occurs very erratically.

16. We also attempted to create an alternative measure, based on a weighted average of the respondents' self-reported attendance rates — the individual weights being proportional to the *inverse* of the individual respondent's attendance rate. However, this proved quite unstable, since respondents with very low attendance rates received extremely large weights.

17. The results printed in Table 3 provide our most conservative estimate of the impact of contributions and resources, based on a subsample of the data that excludes 49 extreme cases. The (more impressive looking) results from the entire sample are:

$$\text{growth} = \underset{(6.75)^{***}}{.218} \cdot \text{contrib} + \underset{(8.19)^{***}}{.265} \cdot \text{attend} \qquad \text{Adjusted } R^2 = .148$$

$$\text{growth} = \underset{(2.07)^{*}}{-.070} \cdot \text{renters} \underset{(5.45)^{***}}{-.200} \cdot \text{yrs\_attend} - \underset{(1.88)}{.061} \cdot \text{members77}$$

$$-.012 \cdot \text{grow7377} + .106 \cdot \text{pop\_growth}$$
$$(0.37)^{*} \quad (2.94)^{**} \qquad \text{Adjusted-}R^2 = .074$$

$$\text{growth} = .244 \cdot \text{contrib} + .223 \cdot \text{attend}$$
$$(7.01)^{***} \quad (6.48)^{***}$$
$$-.109 \cdot \text{renters} - .134 \cdot \text{yrs\_attend} + .042 \cdot \text{members77}$$
$$(3.36)^{***} \quad (3.80)^{***} \quad (1.27)$$
$$.052 \cdot \text{grow7377} + .006 \cdot \text{pop\_growth}$$
$$(1.60) \quad (0.16) \qquad \text{Adjusted } R^2 = .178$$

18. Two sets of questionnaire items ask about church satisfaction. The first set asks respondents whether they are seeking help in nine areas of their lives; the second set asks how helpful the church actually has "been to you in the following areas of your life?" The nine areas were (1) be aware of the needs of others in my community, (2) build good moral foundations for my personal life, (3) find meaning for my personal existence, (4) raise my children properly, (5) know of God's care and love for me, (6) meet my personal problems of anxiety, conflict, etc., (7) strengthen my faith and religious devotion, (8) understand my daily work as Christian vocation, (9) work for justice in my community and my world.

19. When comparing the two sets of variables, one must recall that searches tend to inflate R^2 and overstate significance levels. In other words, it is unlikely that an alternative theory would have predicted, *in advance*, the two or three variables that happen to outpredict resources in this data set. It is also unlikely that the *same* two or three variables will outperform resources in a different data set.

20. The logarithmic regressions corresponding to the linear regressions in Table 3 are as follows:

$$\log(\text{growth}) = .154 \cdot \log(\text{contrib}) + .159 \cdot \log(\text{attend})$$
$$(4.69)^{***} \quad (4.84)^{***} \qquad \text{Adjusted } R^2 = .054$$

$$\log(\text{growth}) = -.121 \cdot \text{renters} - .178 \cdot \text{yrs\_attend}$$
$$(3.91)^{***} \quad (5.48)^{***}$$
$$-.066 \cdot \log(\text{grow7377}) + .134 \cdot \log(\text{pop\_growth})$$
$$(2.20)^{*} \quad (4.13)^{***} \qquad \text{Adjusted } R^2 = .078$$

$$\log(\text{growth}) = .128 \cdot \log(\text{contrib}) + .128 \cdot \log(\text{attend})$$
$$(3.89)^{***} \quad (3.78)^{***}$$
$$-.131 \cdot \text{renters} - .138 \cdot \text{yrs\_attend}$$
$$(3.87)^{***} \quad (3.78)^{***}$$
$$-.014 \cdot \log(\text{grow7377}) + .131 \log(\text{pop\_growth})$$
$$(0.41)^{*} \quad (3.66)^{***} \qquad \text{Adjusted } R^2 = .117$$

The significance levels of the coefficients in these equations mirror those of their linear counterparts. The logarithmic specification might nevertheless be more appropriate given the highly skewed rates of congregational growth and per capita giving found in the UCC data. Log transformations make the distributions of both these variables much more nearly normal.

References

Bromley, David G. 1985. "Financing the Millennium: The Economic Structure of the Unificationist Movement." *Journal for the Scientific Study of Religion* 24:253-75.

Coleman, James S. 1990. *Foundations of Social Theory*. Harvard University Press.

Finke, Roger, and Rodney Stark. 1992. *The Churching of America, 1776-1990: Winners and Losers in America's Religious Economy*. Rutgers University Press.

Greer, Bruce A. 1993. "Strategies for Evangelism and Growth in Three Denominations (1965-1990)." Pp. 87-111 in *Church and Denominational Growth: What Does (and Does Not) Cause Growth or Decline*, edited by David A. Roozen and C. Kirk Hadaway. Abingdon Press.

Hadaway, C. Kirk. 1993. "Is Evangelistic Activity Related to Church Growth?" Pp. 169-87 in *Church and Denominational Growth: What Does (and Does Not) Cause Growth or Decline*, edited by David A. Roozen and C. Kirk Hadaway. Abingdon Press.

Hadaway, C. Kirk, Penny Long Marler, and Mark Chaves. 1993. "What the Polls Don't Show: A Closer Look at Church Attendance." *American Sociological Review* 58:741-52.

Hoge, Dean R., and David A. Roozen. 1979a. *Understanding Church Growth and Decline: 1950-1978*. Pilgrim Press.

____. 1979b. *Technical Appendix to Understanding Church Growth and Decline: 1950-1978*. Hartford Seminary Foundation.

Iannaccone, Laurence R. 1992. "Sacrifice and Stigma: Reducing Free-Riding in Cults, Communes, and other Collectives." *Journal of Political Economy* 100:271-92.

____. 1994. "Why Strict Churches Are Strong." *American Journal of Sociology* 99:1180-1211.

____. 1996. "Reassessing Church Growth: Statistical Pitfalls and Their Consequences." *Journal for the Scientific Study of Religion* 35. Forthcoming.

Jacquet, Constant H., and Alice M. Jones (eds.). 1992. *Yearbook of American and Canadian Churches, 1991*. Abingdon Press.

Kelley, Dean. 1986 [1972]. *Why Conservative Churches Are Growing: A Study in the Sociology of Religion with a New Preface for the ROSE Edition*. Mercer University Press.

Leamer, Edward E. 1978. *Specification Searches: Ad Hoc Inference with Non-experimental Data*. Wiley.

Lofland, John. 1979. "White-hot Mobilization: Strategies of a Millenarian Movement." Pp. 157-66 in *The Dynamics of Social Movements*, edited by Mayer Zald and John McCarthy. Winthrop.

McAdam, Doug, John D. McCarthy, and Mayer N. Zald. 1988. "Social Movements." Pp. 695-737 in *The Handbook of Sociology*, edited by Neil J. Smelser. Sage.

McCarthy, John D., and Mayer N. Zald. 1977. "Resource Mobilization and Social Movements: A Partial Theory." *American Journal of Sociology* 82:1212-41.

McCarthy, John D., and Mark Wolfson. 1993 "'You Can Make a Difference': The Role of Agency, Strategy, and Structure in Grass-Roots Resource Mobilization." Unpublished paper.

Meyers, Eleanor, and Daniel V.A. Olson. 1991. "A Contemporary Profile of the Christian (Disciples of Christ) Church." Pp. 509-20 in *A Case Study of Mainstream Protestantism: The Disciples' Relation to American Culture 1890-1989*, edited by Newell D. Williams. Eerdmans.

Oliver, Pamela E., and Gerald Marwell. 1992. "Mobilizing Technologies for Collective Action." Pp. 251-72 in *Frontiers in Social Theory*, edited by D. Morris and Carol McClurg Mueller. Yale University Press.

Olson, Daniel V.A. 1987. *Networks of Religious Belonging in Five Baptist Congregations*. Ph.D. diss., Department of Sociology, University of Chicago.

____. 1989. "Church Friendships: Boon or Barrier to Church Growth." *Journal for the Scientific Study of Religion* 28:432-47.

____. 1993. "Congregational Growth and Decline in Indiana among Five Mainline Denominations." Pp. 208-24 in *Church and Denominational Growth: What Does (and Does Not) Cause Growth or Decline,* edited by David A. Roozen and C. Kirk Hadaway. Abingdon Press.

Robbins, Thomas. 1988. *Cults, Converts, and Charisma: The Sociology of New Religious Movements.* Sage.

Roof, Wade Clark, and William McKinney. 1987. *American Mainline Religion: Its Changing Shape and Future.* Rutgers University Press.

Roozen, David A., and C. Kirk Hadaway (eds.). 1993. *Church and Denominational Growth: What Does (and Does Not) Cause Growth or Decline.* Abingdon Press.

Royle, Marjorie H. 1993. "The Effect of a Church Growth Strategy on United Church of Christ Congregations." Pp. 155-68 in *Church and Denominational Growth: What Does (and Does Not) Cause Growth or Decline,* edited by David A. Roozen and C. Kirk Hadaway. Abingdon Press.

Stark, Rodney, and Charles Glock. 1968. *American Piety.* University of California Press.

Warner, R. Stephen. 1993. "Work in Progress toward a New Paradigm for the Sociological Study of Religion in the United States." *American Journal of Sociology* 98:1044-93.

Book Reviews

Japanese Auto Transplants in the Heartland: Corporatism and Community.
By Robert Perrucci. Aldine de Gruyter, 1994. 186 pp. Cloth, $37.95; paper, $18.95.

Reviewer: CHOONG SOON KIM, *University of Tennessee at Martin*

In this case study, Perrucci delineates portraits of six Japanese automobile transplants (Honda, Nissan, Mazda, Diamond Star, Toyota, and Subaru-Isuzu) in the Midwest corridor, including the states of Michigan, Illinois, Indiana, Ohio, Kentucky, and Tennessee. After examining "the actions of corporations, political officials, workers, environmentalists, and business people involved in the transplant project as reflections of opposed interests of contending social groups and classes," he concludes that "economic development policies and programs in a local community are shaped by the competing interests of social groups and classes." This book is essentially a study in political economy.

Under the theoretical orientation of so-called embedded corporatism, the author searches for answers to four major questions: Why did the transplants locate in the Midwest corridor? How is a state's economic incentive package presented to various interest groups and the public? How will U.S. workers respond to new approaches to developing worker commitment and to the new strategies of changing the organization of work at the point of production? What will be the long-term impact of the auto transplants on the local and regional social and economic structure?

In an appendix, Perrucci explains his various data-gathering techniques. His inclusive method makes this book a rich source of information and makes it more likely that the volume will attract a wide readership outside the boundaries of sociological circles. One of its strengths is "thick description" — to borrow the words of Clifford Geertz — that relates the voices of Americans and Japanese expatriates involved in the transplants. The author presents not only an aggregation of statistics and managerial strategies but also a story of a concrete nexus of social relations in a particular industrial setting of the Midwest corridor.

As an anthropologist, however, I sometimes had the impression that Perrucci's interest in generalization tended to override local realities. For example, his use of regression analyses to determine the characteristics associated with the location decisions of the transplants is unconvincing. The outcome of his regression analyses appears to be at odds with the case of Tennessee as I have observed it. Perhaps it may be advisable to repeat Francis L.K. Hsu's warning: "Some of our scientific investigations are not unlike counting, classifying and 'computerizing' pebbles on the beach to determine the causes of rising and ebbing of tides."

Nevertheless, the book's merits and contributions are many, and its shortcomings few. It is one of the best works written on the subject and a pleasure to read. I strongly recommend it for advanced undergraduate and graduate students, faculty, and professionals and policymakers at the state level who are involved in foreign direct investment. The book also provides a good up-to-date bibliography.

Places on the Margin.
By Rob Shields. Routledge, 1991. 334 pp. Cloth, $52.50; paper, $17.95.

Reviewer: ROBERT E. PARKER, *University of Nevada, Las Vegas*

This book is intended to serve as a contribution to urbanists, sociologists, anthropologists, and human geographers attempting to explicate the connections between space and culture. Shields is concerned with the everyday use of stereotypes and images of places and spaces, emphasizing that place-names are not simply locations but areas that incorporate a complex set of meanings and images that create social reality. To illustrate his argument, he examines marginal places, areas distinctive in their cultural distance from the norms of the larger society (geographic marginality is less central). The author believes that such places encourage a keener appreciation of a society's core values.

The first part of the book includes an introduction and a theoretical chapter. Shields reveals his ambition in the former when he states his plans to "demonstrate the importance of spatial concepts and categories . . . for the whole way in which we go about thinking about our world." Further, while he notes that his work is exploratory and not designed to provide a new, all-encompassing theory, he claims that his approach to understanding the environment transcends the efforts of those writing (mainly in the 1970s) in the area of semiotics, human geography, and environmental psychology. After an assessment of the merits of the works of Bourdieu, Foucault, and LeFebvre, Shields leans heavily on them, borrowing their conceptual language. He stresses LeFebvre's notion of space, not as an external reality but as a concept that must be understood in terms of the meanings that people ascribe to it. Shields's central theoretical concept, "social spatialization," which refers to the connection between the social, the cultural, and the spatial, though defined in varied ways, is not particularly unusual.

The second part of the book, like the entire volume, is verbose, but unlike the introductory and concluding sections, it is genuinely interesting. It features four case studies in an effort to document that the images and stereotypes of places affect how people understand their worlds. Two places, Niagara Falls and the British seaside resort of Brighton, as well as two spaces (or regions), Northern Canada and the north-south divide in Great Britain, are discussed at length in an effort to illustrate the cultural significance of place images and meanings. Using diaries, films, television, print media, and academic accounts, Shields illuminates the role of the spatial in each of his place studies. While enlightening readers about how, say, Niagara Falls's place image was transformed from "a shrine to the power of Nature," to an "industrial powerhouse," to "honeymoon capital of the world," Shields fails to anchor his theoretical discussion adequately. Indeed, some key theoreticians, such as anthropologist Victor Turner, whose ideas Shields credits heavily in the Niagara Falls study, are not included in the theoretical introduction. More generally, although the studies adequately illustrate the sociospatial approach, they suffer from not being linked to a cultural theory that clearly shows the causal connections between the creation of place images and the actions of people.

The concluding section, a chapter entitled "Synthesis and Implications," falls short of convincing the reader that the text transcends contemporary sociospatial analyses. The strength of the book lies in Shields's use of case studies to demonstrate the import of the spatial in social practices. But the introduction and the concluding

section are somewhat confused, not well connected with the examples, and offer little to scholars familiar with the area. Minus the theoretical packaging, the case studies would warrant reading by novices and others less familiar with the socio-spatial perspective.

The Politics of Migration Policies: Settlement and Integration — The First World into the 1990s.
Edited by Daniel Kubat. Center for Migration Studies, 1993. 379 pp. $14.95.

Reviewer: OTIS L. GRAHAM JR., *University of California, Santa Barbara*

The first edition of this useful book appeared in 1979, exploring "surprisingly parallel" experiences of the nations of the First World (the industrialized countries of Europe, North America, and Oceania) with rising levels of emigration from the Third World in the 1960s and 1970s. That volume found that each of the 19 nations surveyed, from Austria to Australia, had been forced by incessant waves of south-north migration to "realize that in-migrations may have lost their justification." It concluded that concerns over emigration costs prevailed over traditional emigrant-welcoming attitudes and that the "in-migration overrun" of the countries of northwestern Europe was now checked by a turn to restrictive policies. As for the "countries supplying the migrants," they "have been forced to see that outmigration does not solve, but only postpones those problems which prompted it."

These words have an odd ring today, much as would any 1979 report from the Pentagon or CIA on the economic and political health of the menacing Soviet Union. Sending countries still send their millions of migrants, but at higher rates, and their governments apparently continue to postpone problems through out-migration rather than face them. This new, 1993 edition of *The Politics of Migration Policies* also reveals that most emigrant-receiving First World societies (Japan is an exception, and the picture in the vast territories of the former Soviet Union is formidably confused) continued through the 1980s and into the 1990s to absorb heavy waves of migration. The turning point forecast in the 1979 edition did not occur, apart from the termination of guest-worker programs; massive south-north migration accelerated, augmented by the dissolution of the Soviet Union. So much for prediction in the social sciences.

The essays in the new edition, many revised by their original authors, thus extend the story of south-north migration pressures to the threshold of the 1990s. The parallelism of emigration experience among First World nations has expanded beyond the inflow of rising numbers to include an almost universal increase in restrictionist sentiment. (The anthology misses the most intense period of emigration restriction politics that began in Western Europe and the U.S., especially California, in the early 1990s.)

It is useful to have an updated version of the original 1979 essays, along with new essays on Japan and the former Soviet Union. What do the 19 essays all add up to? They do not arrive at any particular conclusion, nor does the editor offer one. In his introduction to the new edition, Kubat ventures some cursory, optimistic thoughts. In view of the fact that the "population on this earth has passed the 5 million mark . . . the migratory flows in Europe seem manageable, which they are." For another flimsy insight, consider this: "A solution is always found and integration and settlement of the foreign population does take place." This bland superficiality

is followed by a sounder observation: "The chief question and concern remaining is to reconcile the political will of the government which errs on the side of benevolent enlightenment and the political will of the people who err on the side of dislike of foreigners' habits both at work and on the street and with whom they compete." Kubat is right that after three decades of emigration pressures on the First World, governments are out of touch with the anger of unconsulted citizens. As for the history-based consolation that "a solution is always found," Kubat's readers will want to consult Robert Kaplan's "The Coming Anarchy" in the February 1994 issue of the *Atlantic Monthly*. The future, Kaplan cogently argues, is going to be quite different from the past.

A volume coming out of the emigrant-advocacy Center for Migration Studies naturally votes for the benevolent acceptance of whatever international population movements seem under way. Although such is the thrust of this volume, readers are free to form their own conclusions from this conclusion-less but panoramic tour of one of the key dynamics of contemporary and future history.

Organizing, Role Enactment, and Disaster: A Structural Theory.
By Gary A. Kreps and Susan Lovegren Bosworth, with Jennifer A. Mooney, Stephen T. Russell, and Kristen A. Myers. University of Delaware Press, 1994. 221 pp. $38.50.

Reviewer: ROBERT A. STALLINGS, *University of Southern California*

Sociologists have been making important contributions to the understanding of disasters for at least a half century, but their research has had relatively little impact on sociology itself. Outside the field of collective behavior, little use has been made of their findings. This neglect is unfortunate, because disasters are natural laboratories wherein properties of social structures and processes are accessible to a degree seldom seen under more normal circumstances.

Gary Kreps, Susan Bosworth, and their research team have been at work for over a decade bringing data gathered in the aftermath of disasters to bear on such important questions about social organization as, What is the essence of organization? By what processes does it emerge, sustain itself, and change? Kreps and Bosworth's new book provides both a review of their earlier findings and a report on the most recent phase of their work, the latter of which involves building a formal theory of role enactment in disasters.

Kreps and Bosworth break down role enactment into three components. *Status-role nexus* refers to whether roles enacted by participants in organized disaster responses are consistent with their predisaster statuses. *Role linkages* identify whether postdisaster relationships reflect continuity with predisaster role sets (e.g., doctor and nurse) or discontinuity (e.g., a college professor and a meat inspector who form a search-and-rescue team). *Role performance* identifies various degrees of innovativeness both by incumbents enacting predetermined roles (role playing) and by those entering into disaster roles that are new to them (role making). Correlation and regression analyses suggest that stability (versus change) in role enactment varies directly with the severity of the disaster, the size of the community in which it occurs, the degree of continuity in the organizational structure, and the speed with which individuals become involved in the disaster response. Furthermore, stability varies inversely with the complexity of the division of labor in the organized response, the

extent to which postdisaster roles involve leadership, and the presence of previous community experience with disaster.

A methodological contribution of this project is its coding of a unique archive of qualitative data. Now housed at the University of Delaware's Disaster Research Center, the archive contains nearly 6,000 in-depth interviews gathered over 30 years on purposive samples of organizations and groups. These are unstructured interviews in which interviewees describe (as respondents) their own behavior during a disaster and (as key informants) characteristics of the organized responses in which they participated. Kreps and his team have coded over 1,000 interviews describing 423 organized responses. A nonrandom subsample of 257 individuals participating in these responses is used in most of the statistical analyses.

Those unfamiliar with Kreps's project will find this book difficult to read, not because of the writing (it is well written) but because the authors have been immersed in their work for so long that the reader will have to work hard to keep up with them. Those familiar with organization or role theories may flinch at some of the authors' conceptualizations and operationalizations. And the empirical work on role dynamics may ultimately uncover little that was not already suspected (e.g., role enactment is more stable within organizational forms that themselves display greater continuity).

Nevertheless, this is a book that sociologists in general — not just those already engaged in research on disasters — will find useful. Kreps, Bosworth, and their student collaborators document issues and steps involved in theory construction, present a formal theory of the dynamics of organizing and role enactment, and provide a detailed protocol to guide future efforts to test their theory with primary data. The book also contains a detailed accounting of successes and frustrations in converting existing qualitative materials into quantitative variables. Most important, the authors provide a primer on how the seemingly esoteric topic of disasters represents an excellent opportunity to explore fundamental questions of organization and organizing.

Participant Observer: An Autobiography.
By William Foote Whyte. ILR Press, 1994. 346 pp. Cloth, $52.00; paper, $24.95.

Reviewers: PATRICIA A. ADLER, *University of Colorado;* PETER ADLER, *University of Denver*

The author of our most enduring and widely read ethnography treats us to a first-person account of his life and times, an elegantly written and carefully documented portrait of the consummate scholar. Covering more than 50 years of his life in academia, Whyte takes his reader on a journey through three continents, four universities, and uncountable research sites. We learn about his family life, the influence of pivotal colleagues, and the effects of a near-debilitating battle with polio. All of this is played against the backdrop of the times in which he lived: before World War II he conducted the research for *Street Corner Society*; in wartime Oklahoma he made a brief, albeit important, foray into the oil industry; in the postwar period he found fertile ground for organizational studies in the manufacturing region surrounding Ithaca; in the tumultuous 1960s he served as a voice of reason in a Cornell racked by violent protest; and in the 1960s through the 1980s he honed his singular brand of applied sociology by integrating research, consultancy, and advocacy in community and organizational studies in Venezuela, Peru, and

Spain. Whyte's stories are told in brief vignettes, with each of the 42 chapters outlining a specific event or crucial epiphany in his life. While traversing this terrain, he offers a history of American sociology. The book is well written, with an eye for developing the relationship between personal biography and academic pursuits.

Though Whyte has not embraced some of the recent postmodern turns in the practice of ethnography, his autobiography can be read with a 1990s sensibility. We are reminded that, as revolutionary as the current practice of ethnography may feel, this important figure sowed the seeds for these changes long ago. As a "confessional" tale, the book poignantly and candidly recounts some of the foibles, drawbacks, and elation that ethnographers face during the course of a research project. Chapter 11, entitled "Learning to Be a Participant-Observer," is a veritable primer for the novice ethnographer on how to gain entrée to research settings. The book also reflects the current fascination with the role of self in qualitative research, exploring how the personal intermixes with the professional and how subjectivity operates to allow ethnographers greater access to their subjects. We also see aspects of the battle over voice and authority, where researchers have to grapple with how much they can speak for their subjects and how much they must allow indigenous populations to speak for themselves. Finally, Whyte's own contribution to ethnography — participatory action research — presaged the present-day focus on action and praxis in the conduct of research.

William Foote Whyte has provided another enduring document for future generations to read and ponder. Those who have already read his other personal accounts in the appendixes to the editions of *Street Corner Society* and his previous work *Learning from the Field* will find that their appetites have been only whetted for this thorough treatment. Though Whyte is best known for one book, his autobiography reminds us of how wide-ranging his contributions have been throughout his distinguished career (he has written or edited eight books since his retirement in 1979). This handsome volume, complete with priceless photographs, should sit proudly on any sociologist's shelf. It will serve as a motivation for productivity, creativity, and commitment to the academic enterprise.

Secularization, Rationalism and Sectarianism: Essays in Honour of Bryan R. Wilson. *Edited by Eileen Barker, James A. Beckford, and Karel Dobbelaere.* Clarendon Press, 1993. 322 pp. $55.00.

Reviewer: JOSÉ CASANOVA, *New School for Social Research*

This is a compilation of 16 essays written by eminent scholars in the social scientific study of religion. What distinguishes this well-deserved and rewarding Festschrift, giving it an unusual coherence for this kind of publication, is the fact that besides having been written by a group of colleagues who share "respect, admiration, and affection" for Bryan Wilson, the doyen of sociological studies of religion in Britain, the essays were specially commissioned for the book and revolve around the three concepts that have served as leitmotifs in Bryan Wilson's rich and influential lifework: secularization, rationalism, and sectarianism.

Some of the essays are only indirectly related to Wilson's work through thematic affinity. These include James A. Beckford's comparative discussion of the different patterns of governmental management of controversial new religious movements (cults) in the U.S. and the United Kingdom; Jean Séguy's historical reconstruction of

the apocalyptic theme in "Adventist" religious orders; Philip Rieff's literate and cryptic discourse, concatenated into 40 epigrammatic theses on the politics of culture and the culture of politics; and Asa Briggs's discussion of British religious broadcasting.

The essays by Karel Dobbelaere, Sabino Acquaviva, and Philips Hammond and Mark Sibley review the old paradigm of secularization in view of the numerous recent criticisms coming primarily from students of religion in America. From diverse perspectives, the essayists all support Wilson's defense of an unreconstructed paradigm that tends to treat the exceptional European case as a universal process of modernization. But this new round in the secularization debate only restates old platitudes. The main fallacy in the old paradigm, reproduced by apologists and critics alike, is the failure to distinguish analytically among three distinctly different, uneven, and unintegrated propositions: secularization as differentiation of the secular spheres from religious institutions and norms (which remains the valid core of the theory of secularization); secularization as decline of religious beliefs and practices (which remains an empirically valid proposition for some modern societies [e.g., Europe] but is patently invalid for others [e.g., the U.S.]); and secularization as marginalization of religion to a privatized sphere (which is a historical option for religions in the modern world). Only by examining and testing the validity of each of these propositions independently of each other can one hope to leave behind the fruitless secularization debate.

The rest of the essays take some of the central categories of Wilson's work and offer various reformulations of them, examine them critically in view of newer research, or apply them to new areas. David Martin, Michael Hill, and Susumu Shimazono investigate Wilsonian patterns of sectarianism respectively in Latin America, New Zealand, and Japan. Roy Wallis, Eileen Barker, and Richard Fenn examine uses of the concept of charisma and the dynamics of charismatization and domestication. The Africanist J.D.Y. Peel revisits *Magic and the Millennium* and returns convinced of the continuing relevance of the Wilsonian framework for the study of West African religion. Indeed, he asserts the analytical superiority of Wilson's typology of religious movements over the usually more contextualized but untheorized Africanist analyses. The most illuminating essays are by Roland Robertson and Richard Gombrich. Both offer a deconstruction of the essentialist category of "religion" as a modern Western construct. Robertson traces the genealogy of the category of religion, its discovery and uses by Western modernity, and its global diffusion and paradoxical indigenization in non-Western societies. Through a comparison of Sinhalese and Japanese Buddhisms, Gombrich traces the dissociation of secularization and Protestantization, two processes that appear intrinsically related in the Weberian-Wilsonian theory of rationalization. Overall, the Festschrift honors the complexity and widespread influence of Wilson's work and meets the expectations raised by bringing together such a roster of prominent scholars of religion.

Public Reactions to Nuclear Waste: Citizens' Views on Repository Siting.
Edited by Riley Dunlap, Michael E. Kraft, and Eugene A. Rosa. Duke University Press, 1993. 332 pp. Cloth, $49.95; paper, $24.95.

Reviewer: WILLIAM M. BABIUCH, *National Renewable Energy Laboratory*

Occasionally a book comes along that is "required reading" in a particular field. *Public Reactions to Nuclear Waste* is one such book. It contains a variety of essays that provide an in-depth examination of the social and political factors associated with the disposal of high-level nuclear waste. The empirical data will be valuable to both social scientists and policymakers as they seek a better understanding of the factors affecting public attitudes toward nuclear issues in hopes of efficiently and effectively resolving America's nuclear disposal problem.

The essays in this volume address broad-based issues related to nuclear waste, such as the context of public concern with nuclear waste and public reactions to initial attempts of the Department of Energy (DOE) to locate repository sites, as well as public opinion issues specific to the site at Yucca Mountain, Nevada. The authors use a broad range of data-collection techniques and provide valuable descriptions of their methodologies to allow the reader to understand the strengths and weaknesses of their findings.

While each essay addresses the issue from a slightly different angle, two key findings appear throughout the book. First, a high-level nuclear waste repository is unacceptable to most Americans for the foreseeable future because the associated risks are perceived to be too great. As noted in an essay by Desvousges, Kunreuther, Slovic, and Rosa, "Although [the respondents] show recognition of the economic benefits associated with the repository, [they] do not believe those outweigh the risks." Second, perceptions of the risks of a high-level nuclear waste repository are closely tied to trust in government and in agencies responsible for the repository management, especially DOE. This book strongly contends that restoring public trust in DOE's handling of nuclear matters is vital if America hopes to deal effectively with nuclear waste. The importance of trust has not gone unnoticed by DOE. In April 1991 Secretary of Energy Watkins created the Task Force on Radioactive Waste Management to recommend measures that DOE might take to strengthen public trust and confidence in the civilian nuclear waste management program. The report issued by the task force states: "It is the strong view of the Task Force that DOE stands little chance of strengthening public trust and confidence unless it recognizes that its decision-making behavior will have to fundamentally change." The report goes on to call for greater inclusion and empowerment of parties external to DOE in the decision-making process. Although DOE is making an effort to restore public trust and confidence, findings presented here confirm that it will be a difficult task, for "it is much easier to lose people's trust than regain it."

In combination with sound empirical analysis of the public's attitude toward nuclear waste, this book examines policy prospects for the future of nuclear waste. In the closing essay, Rosa, Dunlap, and Kraft assess the value of "traditional fixes" and express concern that each of these fixes minimizes direct public participation. Such minimization runs counter to recommendations made in the DOE task force report as well as in this volume. Kraft and Clary contend, for example, that even though public hearings are one way to obtain public input, more meaningful public involvement is necessary if DOE hopes to restore its credibility and find a publicly acceptable solution to the problem of nuclear waste.

In my work as an energy analyst I have heard DOE policymakers stress that they "need substantive data to support claims about the public's perception of energy-related issues." This volume meets that need, providing data on the public's opinion of nuclear energy in general and nuclear waste in particular. The book will be an invaluable contribution to the resolution of America's nuclear waste disposal and should be read by all social scientists and policymakers involved in this issue.

Continual Permutations of Action.
By Anselm Strauss. Aldine de Gruyter, 1993. 280 pp. Cloth, $44.95; paper, $21.95.

Reviewer: GARY ALAN FINE, *University of Georgia*

Perhaps no sociologist of the past half century — with the possible exceptions of Erving Goffman and Robert Merton — has been as generous in creating a string of provocative middle-range concepts, given freely to the sociological community, as has Anselm Strauss. In his latest contribution, forgettably entitled *Continual Permutations of Action*, he sets out to develop a "theory of action" — that holiest, elusive grail of sociological endeavor, the Unified Field Theory. If the field is not so easily unified and action not so easily known, it is not for want of worthy attempts.

In fairness, the book should not be considered a grandly synthetic work but rather an important summation of the relations among the constructs that Strauss has developed over his long career. It is a return to certain basic concerns that he addressed in his important 1959 volume, *Mirrors and Masks*, and that have animated his career. (It is astounding to recognize that *Mirrors and Masks* was a midcareer contribution, and if his productivity proceeds apace, *Continual Permutations of Action* will be as well.) In *Mirrors and Masks*, Strauss argues, probably more effectively, if with less bravado, than does Goffman in *Presentation of Self*, that actors are lodged within a nexus of interaction and structure.

Action consists of the possibilities of linkage between the individual and a variety of social structures as mediated through situations. Strauss's theoretical perspective, which emphasizes the centrality of collective action, is in direct line with the pragmatism of Dewey and Mead: the focus is on the ongoing components of *acting*. Strauss underlines the contingent nature of social order in collective action, a processual view that has posed barriers to the creation of an interactionist version of social organization. How can organization be powerful in a world of contingencies? Strauss's particular contribution is to outline the bridge between an actor's choices and the reality of structure.

Most of Strauss's numerous concepts — among them, the arc of work, negotiated order, awareness contexts, and policy arenas — share a creative tension between acting and constraint. The reality of the other provides an arena in which meaning is created and interaction generated. This model is distinct from unmoored visions that assume the nearly unbridled creative power of the mind. Strauss gives primacy to neither cognition (strong social constructionism) nor affect (certain strains of postmodernism). In contrast, the primacy of acting reminds us that this grand theory is a branch of social behaviorism — behavior that is limited, contextualized, constrained.

Unlike Strauss's rich treatment of action, context, and organization, his theory of self is markedly attenuated. If, as has been said, interactionists come in two flavors, self-theorists and situation-theorists, Strauss is of the latter breed. The body

has more weight than the soul. Whereas Blumer keeps searching for the self — the core, I — Strauss hardly notices its absence as long as a body is present to act and a biography remains as an evidentiary record to be employed in the creation of relations.

Strauss begins *Continual Permutations of Actions* by modestly acknowledging that he did not consider himself to be a social theorist until colleagues urged him to make his model less implicit. As one colleague is quoted as saying, "Some of your best ideas are in your subordinate clauses." The reader may find this oddly insightful comment simultaneously valedictory and critical. One completes the book without having confronted a grand theory but having had numerous — perhaps too many — encounters with ideas and images. (For evidence of this cornucopia, compare Blumer's *three* premises of symbolic interaction with Strauss's *19* assumptions.) This volume, which explicitly attempts to generate a theory of action, can most properly be said to provide the tools by which an integrated model *might* develop. Too much is provided to be considered parsimonious, but just enough to allow us to glimpse a set of complex truths.

In the end this book proves not to be the simplified, explicit "theory" that Strauss's colleagues had wished, but it represents Anselm Strauss at his theoretical prime: the master of the sensitizing concept.

Editors as Gatekeepers: Getting Published in the Social Sciences.
Edited by Rita J. Simon and James J. Fyfe. Rowman & Littlefield, 1994. 275 pp. Cloth, $53.50; paper, $22.95.

Reviewer: MICHAEL SCHWALBE, *North Carolina State University*

Suppose we wanted to know how corporate elites keep their boardrooms looking so white and male. One way to find out would be to ask them. A likely response might be, "We don't discriminate. We look for qualified people who are team players, people we feel comfortable with and can trust to do what's best for the company. Gender and race have nothing to do with it." After collecting many such accounts from articulate, polite, and sincere executives, it might begin to seem that charges of racial and gender bias are nothing but ego balm for sore losers.

Suppose we wanted to know how editors make decisions. One way to find out would be to ask them. If we put their answers into a book, the result would look like Simon and Fyfe's *Editors as Gatekeepers.* In this book, editors of journals in various disciplines and editors of commercial and university presses give accounts of what they do, how, and why. But like the accounts we might get from corporate elites about how they reproduce themselves, these are more valuable as data than analysis.

This is not to say that the accounts in this book are uninteresting or ring false. In fact, the authors seem to be trying hard to offer an honest look backstage. There are even some surprisingly frank confessions.

For example, former *American Sociological Review* gatekeeper Rita Simon says, in her contribution to the volume, that an editor who dislikes a paper that has nonetheless received two positive reviews can send it out again, into the jaws of tougher critics, who are likely to provide justification for the decision toward which the editor is inclined. To us commonfolk this might seem unfair, but as Simon

explains, editors are people of extraordinary talent and accomplishment, and thus deserve the freedom to maneuver in these ways.

In other chapters we learn that some journal editors prefer to publish papers that give their disciplines the appearance of coherence; that a paper for which an editor cannot quickly think of a reviewer is probably doomed; that for some editors the first step in the review process is to identify the author as a bigshot, master craftsman, junior apprentice, or crank; that the review process treats everyone alike; that apostates of positivism are unlikely to be chosen as editors; and that most complaints of bias on the part of rejected authors are unfounded since, as these editors attest, the system works pretty darn well.

These tidbits suggest some of the contradictions within and between the book's chapters. There are others. For example, peer review, we are told, encourages writers to perfect their papers before sending them out; yet this flies in the face of the 90 percent rejection rates at top journals. We are also told that editors use rigorous, universalistic standards for evaluating papers; yet this contradicts other confessions about how the process is manipulated to get the results editors want. And then we are told, over and again, that the review process is healthy, even though 41 percent of sociologists say that the articles in major journals are boring and even though about 70 percent of those articles are never cited.

It would have been nice if Simon and Fyfe had taken these contradictions as starting points for a more critical analysis of how editors tend their gates and what the results are. It would also have been nice to see the chapter authors, or Simon and Fyfe, interrogate their ideas about what constitutes an important, useful, original, and insightful contribution — as many speak of looking for. What is considered important and useful? Why?

When corporate elites tend their gates, they give similarly vague accounts of the criteria they use to decide who gets in. Those accounts may be revealing in some ways, but they are not *explanations* of what really goes on. We should not expect otherwise. Gatekeepers are, by definition, committed to reproducing, not critically analyzing, the communities that put them at the gates. To understand more deeply what they do and why, it is necessary to examine what gatekeepers take for granted, much of which remains taken for granted in this book.

Publishing junkies, gripers, cynics, and curious graduate students will probably enjoy *Editors as Gatekeepers*. Some of the essays are a bit clunky and annoyingly self-congratulatory (the gate should have been narrowed, let's say), but on the whole they are easy to read. I doubt that the book will really save neophytes much "time and pain," as Simon and Fyfe claim. But at least it will give them a better idea of the grief they're in for.

George Herbert Mead: The Making of a Social Pragmatist.
By Gary A. Cook. University of Illinois Press, 1993. 240 pp. Cloth, $31.95; paper, $15.95.

Reviewer: HOWARD L. SACKS, *Kenyon College*

Over the last decade, the work of George Herbert Mead has been a subject of renewed interest and critical examination. Scholars have linked Mead's thought to European sociological and philosophical traditions and have expanded the available body of Mead's work through the publication of previously unpublished essays and

his students' lecture notes. Taken together, this scholarship has reestablished Mead as a seminal American thinker whose ideas have contemporary significance for sociology. In part, recent scholars have accomplished this task by wresting the Meadian legacy from the grasp of symbolic interactionism, which has too often distilled and trivialized Mead through a narrow reading of only a portion of his writings.

Gary Cook's intellectual biography contributes to this expanded view with a careful textual examination of Mead's book reviews, scholarly essays, and writings associated with his social reform activities in Chicago throughout the first decades of this century. Cook begins, appropriately, with Mead's early development at Oberlin College and his philosophical studies at Harvard and in Germany. Here we learn of the intellectual dilemmas that engaged Mead and his close friend Henry Castle and that would animate Mead's later scholarly and reform activity. From the beginning, Mead sought a rational understanding of human experience that overcame the a priori constraints he saw in both Kantian metaphysics and Christian spirituality. Writing to his friend, Mead disclosed the fear that he had "too much reason and too little strength to be a Christian, and too little reason and too much sentiment to be a philosopher."

Mead found the solution to this dilemma in a dynamic social psychology that remains his most original intellectual contribution. Working closely with his colleague John Dewey, Mead strove to overcome the mentalism and dualism of existing approaches to identity in favor of a functional psychology that emphasized social conduct as the wellspring for defining individual identity and social reality. By grounding identity in social conduct, Cook notes, Mead discovered an alternative to the "assumption that one must opt either for the older conception of consciousness as involving special mental substance or for an outright denial of the existence of consciousness." But although Cook examines the development of Mead's formulation of the social self in great detail, he fails to articulate fully Mead's ideas on the dialectical character of social conduct. Instead, Cook vaguely attributes Mead's dialectical thought to the influence of Hegel, obscuring Mead's more materialistic approach. This is an unfortunate omission, for several reasons. First, the dialectical character of Mead's theory of self, embodied in the concepts of the "I" and the "me," has been a source of considerable debate and confusion. Cook's otherwise painstaking examination of Mead's key ideas might have done much to clarify these central but elusive concepts. Second, by largely ignoring these concepts, Cook repeats the weakness all too common in symbolic interactionism's treatment of Mead, which has contributed to a misunderstanding of Mead's social psychology as consistent with sociological functionalism. Finally, this omission limits our ability to understand the crucial link between Mead's social psychology and his social activism.

Cook makes his most original contribution in his examination of Mead's reform activity and pragmatic vision of social reconstruction. Through their involvement in educational and labor reform and in the settlement of immigrants, Mead and Dewey constantly tested their ideas about society, an empirical moment sociologists have generally overlooked. Cook clearly reveals Mead's valuing of the scientific method as a vehicle for intelligent social change. Nevertheless, Mead rejected a positivist approach that would disallow the scientist's participation in morally grounded efforts at social reconstruction. As Cook so ably demonstrates, in Mead's work, social theory, scientific method, morality, and social praxis are phases of a single

process of constructive action. Taken together, these strands constitute Mead's pragmatist vision.

The serious student of Mead will find few new insights here, but the sociologist with a more casual understanding of Mead's work will gain a deeper sense of Mead's key ideas from this ably argued and straightforward presentation. In Cook's portrait of George Herbert Mead we find a morally engaged teacher and scholar whose intellectual life is inextricably bound to the betterment of his community. It is a model that sociologists, too often removed from the world they seek to understand, would do well to emulate.

Simulating Societies: The Computer Simulation of Social Phenomena.
Edited by Nigel Gilbert and Jim Doran. UCL Press, 1994. 306 pp. Cloth, $59.95.

Reviewer: KENNETH A. BOLLEN, *University of North Carolina at Chapel Hill*

Simulating Societies grows out of an international symposium of the same name that was held in 1992 at the University of Surrey, Guildford, England. The book has 12 papers by 23 contributors. Doran and Gilbert's introductory chapter provides a definition and overview of simulation models. They use the collapse of Mayan society to illustrate some of the ideas. Despite the concrete example, the presentation tends to the abstract. Séror's chapter provides an overview of different simulation strategies and their epistemological dimensions. Troitzsch's chapter is more concrete. He examines the evolution of hunting and gathering, agrarian and industrial technologies in fragmented populations. Nowak and Latané describe a simulation program that explains how the attitudes in a population can change due to the strength and immediacy of the attitudes and the number of people holding different or similar attitudes. The chapter by Penn and Dalton discusses urban traffic but has implications for the broader understanding of spatial factors on human behavior. The details and data of this chapter exceed those of the earlier ones. An ant colony simulation is the subject of Drogoul and Ferber's chapter. Of most relevance to sociologists is their discussion of a method where macro behavior is built from micro behavior.

A chapter by Bousquet et al. considers the interaction between people and environmental resources in a society of fishermen while Mithen's chapter simulates prehistoric hunter-gatherer societies. Mithen's review of the major simulation approaches to hunting and gathering societies is quite informative. The chapter by Doran, Palmer, Gilbert, and Mellars is a progress report on a large simulation for modeling upper Palaeolithic social change. Reynolds uses what he terms *cultural algorithms* to study the microlevel evolution of individual traits and macro-evolutionary changes in beliefs. They use this approach to simulate the rise of cooperation among llama herders on the punas of Ayacucho, Peru.

José Castro Caldas and Helder Coelho have one of the most interesting chapters in the volume. They examine bounded rationality and oligopoly among economic agents in markets. I particularly liked their use of human experiments that they analyzed prior to replacing the human agents with simulation ones. Such an approach gives greater validity and realism than is typical in simulation models. Conte and Castelfranchi's concluding chapter is a critique of artificial intelligence models of social action. More specifically, they question the assumptions of agents' cooperation, full information, and rationality.

The collection of chapters may be viewed as a social scientific sampler of simulation models. Much of the work comes from anthropology, social psychology, and economics. Though Doran, a coeditor, is a sociologist, the influences of these other social sciences far exceed those of sociology. This is more a reflection of the relative neglect of simulation research in sociology than a failing of the book. Indeed reading this collection of papers made me wonder why simulation techniques are not more widespread in sociology. Undoubtedly, the oversimplistic simulation models from the 1960s and 1970s left many sociologists skeptical of such an enterprise. Also, the small percentage of sociologists trained in the mathematical and computer programming skills that are required for such efforts limits the pool of researchers capable of conducting such simulations. Nevertheless, several papers in the volume examine issues of intense current interests in our discipline — e.g., the construction of macro structures from the behavior of individuals, the intersection of sociology and economics, the linkages of society and environment. Perhaps it is time for sociologists to take another look at simulations. This edited volume is a reasonable starting point to sample projects from neighboring disciplines.

African Americans: Essential Perspectives.
Edited by Wornie Reed. Auburn House, 1993. 160 pp. Cloth, $49.95; paper, $16.95.

Reviewer: MARLESE DURR, *Wright State University*

The four original essays in this book suggest that social, economic, political, and judicial issues in the African American life cycle require compacted analysis. An underlying theme is that small societal changes since the 1960s have recast African Americans' social status, but structural and economic changes may have slowed their further progress. A second cleverly veiled theme presents African American value orientations as by-products of our racial stratification system. The first two essays relate specific historical parallels to what the authors consider to be the current condition of African Americans — perseverance despite small-scale social change.

The third essay addresses structural changes within the economy, giving special emphasis to labor market reconstruction, which the authors contend is now managerially based. The critique implies that the transition in authority structure from capitalists to managers and policymakers marginalizes African American managers and policymakers, who mostly supervise the African American working class. This raises the questions of ways that African Americans can act independently of the "public sector" helping positions that have historically been their role as public servants, and how they can gain admission to newer occupations.

Acknowledging that African American political participation was and is a search for power, the authors advise persistent use of traditional and nontraditional methods of dissent to realize the goal of social and political parity. The highlight for readers here is the speculation about the growing number of African American conservatives. The authors explicitly claim the presence of a strong liberal orientation among African American conservatives, while subtly positing that this information might have been known to social science researchers had they made a systematic study of political attitudes for this social group.

The final essay underscores race as a significant factor in the examination of issues of justice. Discussing sentencing, capital punishment, juvenile justice, and

drugs, the authors suggest four issues that need scholarly investigation: (1) discrimination in the plea bargaining process; (2) different treatment of African American female and male defendants; (3) the importance of the race of the victim; and (4) discrimination in the formal parts of the criminal justice process. In addition, these authors note the paucity of investigation regarding the decision to reject bail and the dismissal of charges once filed. Citing ambiguous and contradictory data on minority youth crime, the authors briefly discuss the discretionary power of the juvenile justice system and its severe impact on employment opportunities for African American youth upon release.

These essays may not provide new data, but their historical comparisons offer food for thought and suggestions for further research.

The Classic Roots of Ethnomethodology: Durkheim, Weber, and Garfinkel.
By Richard A. Hilbert. University of North Carolina Press, 1992. 260 pp. $39.95.

Reviewer: JAMES J. DOWD, *University of Georgia*

It is interesting to consider the corrosive and redefining properties of time and history. The challenge to mainstream sociology posed by the ethnomethodological "movement," as some (see Pierce J. Flynn's *The Ethnomethodological Movement* 1991) consider the writing and research of Garfinkel, Cicourel, Sacks and others to be, no longer seems quite so interesting. In fact, in the wake of the postmodernist surge, this challenge seems remarkably conservative in its insistence on the precise analysis of carefully coded conversations and other discourse. Hilbert's analysis of the elective affinities between ethnomethodology and the writings of Durkheim and Weber has much to recommend it but also suffers from a lack of originality that, given the author's focus on the arcane niceties of whether Parsons or Garfinkle provides a better reading of Durkheim and Weber, consigns this book to the farther reaches of the shelf.

The central themes of Hilbert's argument have already been developed in John Heritage's *Garfinkle and Ethnomethodology* (1984), which Hilbert cites with admiration. Hilbert's thesis is that Parsons, while justifiably important for his integration of classical writers like Weber and Durkheim, also suppressed some of their essential insights and directly negated some of their other lines of argument. Parsons's student, Garfinkle, in his empirical ethnomethodological studies, reclaims these suppressed and negated themes in Durkheim and Weber. In this sense, ethnomethodology is true to the classics. Ethnomethodology is much closer to Durkheim's original analysis of anomie, for example, than is Parsons, who claimed that anomie leads to behavior breakdown. Ethnomethodology is also much closer to the Weberian emphasis on the empirical study of ideas in history than is Parsons's preference for conceptual work. For Weber, the normative order has no place at all in sociology while, for Parsons, it exists as a second, separate order apart from a factual order. While Hilbert is on solid ground in his critique of Parsons's claim that a normative order exists separately from and prior to the factual order of behavior, the deconstruction of the Parsonsian claim that values drive behavior is not original with Hilbert but has been accomplished by others, most notably by Swidler in her oft-cited paper, "Culture in Action" (1986).

Apart from offering an apologia for ethnomethodology, Hilbert also claims to show that functionalism should be understood as a good instance of Weberian

rational-legal bureaucratization. His intent, then, is not only to provide a strong rationale for ethnomethodology's support of sociology's most productive traditions, but also to demonstrate, as if such demonstration were necessary, that the functionalism of Parsons is flawed. Ironically, however, one is left more sympathetic to Parsons after reading Hilbert than one may have been before. It makes little sense, after all, to insist that because one may gain a new appreciation of Weber or Durkheim after a close reading of Garfinkle, earlier readings of the classics such as, in the present case, Parsons's own classical interpretation from *The Structure of Social Action* must therefore be false.

Neither is Hilbert persuasive in his argument that it has been Durkheim, Weber, and Parsons, rather than the phenomenologists Schutz and Husserl, who have served as the major influences on Garfinkel. The influence of Schutz and Husserl, as well as Parsons, on the sociological imagination of Garfinkle cannot be gainsaid. Not only did Garfinkle attend lectures by Schutz, cite his work continually, and incorporate phenomenological ideas and assumptions into his own work, but his teacher (Parsons) himself is known to have carried on a lengthy correspondence with Schutz.

Toward the end of this book, Hilbert completes his case on behalf of ethnomethodology's importance to sociology by arguing for its compatibility on key issues with other sociological theories such as neofunctionalism, Collins's conflict theory, Giddens's structuration theory, and even Marxism (not to mention Durkheimian and Weberian approaches that had been examined previously). In summary, then, Hilbert offers a provocative and informative discussion of ethnomethodology's roots in the classics, via its encounter with Parsons. Since Garfinkle's project is a radically empirical one, it has never fully developed its theoretical implications or retraced its own theoretical lineage. It is this latter task that Hilbert has successfully accomplished in this work.

Frames of Remembrance: The Dynamics of Collective Memory
By Iwona Irwin-Zarecka. Transaction Publishers, 1994. xiv + 214 pp. $32.95.

Reviewer: JEFFREY K. OLICK, *Columbia University*

On the evidence of this book, Iwona Irwin-Zarecka is a dedicated and good teacher. *Frames of Remembrance* lays out in an accessible, interesting, and thorough manner major methodological and substantive issues in social memory studies. She demonstrates to the novice sociologist as well as to those new to the specialty why we should care about memory. She discusses the multiple forms it may take and roles it may play and gives good advice about how to study it. The book is impressive in the scope of issues it raises and examples it provides, and it is filled with good common sense for those headed into the field.

The book is divided into three major sections. In the first, Irwin-Zarecka justifies a sociological approach to memory and provides a basic orientation drawn on Goffman's frame theory. In the second section, called "Dynamics of Relevance," she reviews various modes, locations, and processes of collective remembrance, including its contribution to identity, its essentially contested nature, its embeddedness in material and popular culture and contemporary politics, as well as its connections with images of the future. The third section focuses more on the social action involved in memory production, including the social organization of

forgetting, memory projects, and the role of opinion leaders in transmitting particular images, interpretations, and constructions of the past.

Throughout the book, the author emphasizes that collective memory is an ongoing process and project. This is a salutary corrective to prevalent presentist approaches in the field — those which see the past either as a tool in social struggles or as their thoroughly malleable outcome. In addition, the author's moral and political concerns are of the sort that are inspiring to students and remind us all of the value of our enterprise. Irwin-Zarecka does not hide her potent worries about the fate of memory, particularly of Holocaust memory, and the Polish case she wrote about in an earlier book forms a backdrop for the entire enterprise.

As a chatty introduction and illustration for undergraduates, and as good advice to graduate students, the book is a worthy contribution. But it is just these strengths, as well as other problems, that prevent the book from making a major contribution to the field. First, the overly conversational tone and digressive writing works in favor of accessibility and relevance but against systematic presentation. This is not wholly unintentional: the author describes the book as "'a meeting place' allowing for an exchange of ideas, concerns, and findings"; the chapters are conceived as "relatively self-contained units, as diverse 'entry points' defined by questions rather than the lines separating disciplines or empirical territories." This strategy, however, seems more suited to asking questions than to answering them. As a result, the presentation is redundant, and divisions seem rather arbitrary. The style and form seem better suited to the classroom than the printed page.

Second, however, these matters have a more substantively troubling programmatic source. The author advocates what she calls a "context-sensitive approach," one that seems to abjure most theoretical and generalizing discourse. She argues that "it may be a long time before there is an analytical agreement on the dynamics of collective memory" and suggests that "as students of the dynamics of collective memory, we may be best advised to stay with the smaller [questions]. . . . The very complexity of the patterns we have found . . . point out the need for scaling down the inquiry." On this point, I strongly disagree. If ever there were a field overrun by unconnected case studies and in desperate need of theory and synthesis, it is social studies of memory. The author provides a useful annotated bibliography at the end, but the book does not come close to being representative, and it does not include some of the most prominent theoretical treatises in the field. Irwin-Zarecka recognizes the dispersion of the enterprise across disciplinary and geographical boundaries and adduces many examples to illustrate problems, but she seems explicitly to avoid synthesis and summary. Indeed, the book lacks a conclusion both formally and substantively.

In sum, *Frames of Remembrance* provides a useful introduction to issues and a good methodological guide, but avoids more definitive gestures, for which I think the field is ready. Nonetheless, if the book is more propadeutic than profound, it is still very much worth a look.

Class Structure and Social Transformation.
By Berch Berberoglu. Praeger, 1994. 122 pp. $45.00.

Reviewer: RANDY HODSON, *Indiana University at Bloomington*

Those interested in an orthodox Marxist interpretation of class structure for their undergraduate courses on social stratification will want to consider this book. It thoroughly develops the idea that class inequality is the central defining characteristic of modern societies.

The first chapter presents a range of theories of social stratification (mainly functionalist, Weberian, and elite theories) and criticizes these from a Marxist standpoint. Marxist class theory is presented next in a straightforward manner by a review of well known material on historical materialism, class as such and class for itself, ideological hegemony, and the role of the state in enforcing class inequality.

Subsequent chapters present a history of class societies and discussions of the current class structures of advanced and Third World societies. The chapter on the history of class societies is perhaps the least strong of these. Too much credit is given to Marx and Engels as historians, and much of the history of class inequality is seen through the lens of their very early understanding of these topics, including their writings on Oriental despotism. The chapter on classes in advanced societies is interesting and readable. The chapter on the Third World is perhaps too detailed. Too many continents, subregions, states, and time periods are reviewed in too brief a period to be easily digested by novice readers.

In the final set of chapters, Berberoglu argues that the understanding of racial inequality, nationalist movements, and gender issues must be based on an analysis of their class content. Berberoglu convincingly argues that national struggles are often motivated by class exploitation between nations. He presents analogous arguments for the class content of race and gender conflicts.

The merits of this book include its brevity, its general accessibility, and its doctrinaire approach. These strengths will make teaching Marxist class analysis from this book relatively straightforward. The content is organized with pedagogy in mind. Students who complain about long readings will appreciate chapters that are on average 12 pages in length (excluding notes). Its merits, however, also constitute limitations. No time or energy is spent considering the limitations of Marx's ideas and concepts that have been revealed by the century and a quarter since his major writings. The decision to present the material in this way may work as a pedagogical device, but an unconsidered presentation of Marxist class analysis is not very intellectually satisfying in today's world and some instructors may have reservations about presenting such ideas uncritically to their classes.

Programmatic statements and questionably grounded visions of nascent proletariat revolutionary movements in both the industrialized and developing worlds may tilt the balance against this book for some instructors. For example, in a section on class struggle in the U.S., the author predicts that: "The limited but real victories achieved in rank-and-file struggles of the past two decades are a beginning for longer-lasting, protracted struggles . . . so that the 1990s (as were the 1930s) would once again become a decade of militant labor struggles at the grassroots level." Similarly, in a discussion of class struggle in the Third World: "The material conditions that have led to U.S. imperialist domination of the world economy up to the present period have now reached a point where broad segments of the masses in the Third World are coming together to challenge it at its very foundations."

On the other hand, for those who like their Marxism straight, such programmatic statements and projected scenarios may be part of the content they want and expect. If this is the case, then this may well be their book; it is loyal to Marx's theoretical tenets and is pedagogically oriented and generally well written.

Parents, Their Children, and Schools.
Edited by Barbara Schneider and James S. Coleman. Westview Press, 1993. 192 pp.

Reviewer: SOPHIA CATSAMBIS, *Queens College, City University New York*

This is the first major publication on the effects of parental involvement in children's education that is based on large-scale survey data. It consists of a series of carefully conducted research papers that use the National Education Longitudinal Study of 1988 (NELS88), a survey of 26,000 eighth graders, parents, teachers, and school administrators. Specifically, it explores how families of different social and economic conditions attempt to further their children's education. It provides empirical evidence supporting the belief that parental involvement improves children's learning opportunities and school performances.

This investigation covers different types of involvements: parental home-based involvement of discussions with children about their educational experiences and plans, supervision of after-school activities, and savings for college; parental community involvement, like contacts with other parents; and parental school involvement, like volunteering at school and participating in parent-teacher organizations.

Some of the study's most significant findings concern race and ethnic variations in patterns of parental involvement. White parents tend to view education as cultural enrichment. They engage in activities that are social in nature, like discussing school with children, knowing their children's friends, and volunteering at school. African-Americans frequently speak with their children about school and are active in parent-teacher organizations. They are the group most engaged in "crisis intervention" activities that involve academic matters, like checking homework and contacting schools. Asian American families expend most of their efforts and resources on out-of-school educational activities. Finally, Latino families have the highest current school expenses.

African American and Latino parents, in comparison to whites and Asian Americans of similar social background characteristics, have the most significant school involvement. The schools that many minorities attend promote high involvement levels by employing minority teachers and by initiating frequent parent contacts. However, although many minority parents seek to compensate for economic or social disadvantages through active parental involvement, they cannot eliminate all of the educational disadvantages that their children face.

Additional findings show that students from all family types benefit from high levels of parental involvement and that various types of involvement affect students' performances differently. Home-based parental involvement mostly affects students' test scores, while school and community-based involvement mostly affects their grades.

Variations in levels of parental involvement partly explain socioeconomic and family-type differences in students' achievements. For instance, parental involvement partly explains the higher performance levels of students from traditional two-parent

families over students from five other family types. Moreover, families in which mothers work part-time have the highest frequencies of all forms of parental involvement.

This study also considers school choice, within the public or private sector, as an indicator of parental involvement. A series of complex analyses reveals that, unlike minority parents, white parents with low levels of education are generally not responsive to opportunities for school choice.

Overall, this book enhances the limited knowledge about parental involvement and identifies important issues. For example, many current policies seek to improve minority achievement by encouraging parental involvement, yet these policies may be based on erroneous assumptions because minority parents are quite active. Therefore, how can parental involvement translate into higher student achievement in schools that serve minority populations?

More research is needed concerning successful parental involvement and how it can be translated into higher achievement levels, increased school engagement, and greater success in all levels of education. NELS88, with its subsequent follow-up surveys of students and parents, will continue to be an essential data source, and the researchers in this series of papers are highly likely to produce future insightful analyses.

Professional Women at Work: Interactions, Tacit Understandings, and the Non-Trivial Nature of Trivia in Bureaucratic Settings.
By Jerry Jacobs. Bergin & Garvey, 1994. 144 pp.

Reviewer: DEBRA RENEE KAUFMAN, *Northeastern University*

The author of this highly readable and compact book touts its importance as one of the few attempts to study work from a qualitative perspective. Jacobs offers *Professional Women at Work* as one way to increase the number of studies offering "rich descriptions of work and the work experience." While I applaud the motivation to move from a positivist emphasis in the study of work to a qualitative one, it is not so clear that ultimately the author delivers the "rich data" the reader expects. Nor do I agree with the strong implication that qualitative studies often fall into the category of the "nontrivial study of trivia." To describe the routine, everyday experiences of respondents on the job as trivia unintentionally supports the notion that there are more profound issues in the sociological study of work. If the study of "trivia" leads to an understanding of how workers come to know and to do their work, then why continue to call it trivia?

I much prefer Jacob's reference to his case studies as "biographies of work" based on interviews with five women: a publications coordinator, an administrative analyst, a child-care administrator, an HIV clinic administrator, and a clinical psychologist. Although Jacobs argues that these are not gender related biographies, closer inspection reveals that, for the most part, the women under study are in the "female sector" of the professional labor force. That is, they are either in middle-management positions within public institutions or in stereotypical nurturing or helping occupations. Gender and race are not randomly distributed in the professional hierarchy. A host of data suggest that ascriptive characteristics do make a difference in how you and your work are perceived, evaluated, and consequently acted upon. By not exploring the gendered context of their experiences in the labor

force, Jacobs misses an important component of the ways in which women negotiate their day-to-day realities. This is particularly true when he discusses their complaints of powerlessness and attributes it primarily to bureaucracy.

Without a detailed description of the context for women's labor market activity, the data lack what Gertz would call "thick description." Except for scattered references about demographics, there are very few examples of how the personal and public lives of these women converge. Similarly, more illustrations would help to buttress the point that the meaning of work is constructed through the innumerable interactions and negotiations that occur on a day-to-day basis. Nonetheless, the book represents a healthy direction for research on women and work in the contemporary U.S.

Circles and Settings: Role Changes of American Women.
By Helena Znaniecka Lopata. SUNY Press, 1994. 325 pp. Cloth $54.50; paper, $17.95.

Reviewer: JENNIFER L. PIERCE, *University of Minnesota*

Helena Zopata's most recent book is a comprehensive overview of the changing roles of American women over the past century. Drawing from her own research, which spans almost 30 years, as well as the extensive literature in the sociology of gender, Lopata employs a feminist symbolic interactionist perspective to examine women's varied roles over historical time and across the life cycle. Her main interest lies in exploring the role clusters women assume (e.g., mother, wife, and paid worker), the resources they have with which to perform these roles, and the level of social and psychological commitment they have to these various roles.

Central to Lopata's analysis is the distinction between traditional, transitional, and modern roles for American women. Traditional roles are enacted in the context of agricultural and pre-industrial societies. She explores these roles in the history of the U.S. as well as other countries. Modern women, on the other hand, are flexible, "highly individuated and competent to function in a complex, industrialized society." While some American women adopt modern roles, Lopata finds that most women today perform their roles in an intermediate or transitional fashion. (In transitional families women assume primary responsibility for raising children.) She examines the diversity of women's experiences not only in terms of these ideal types, but also within the context of social class, and racial and ethnic differences.

The book is divided into eight chapters. Chapter 1 reviews role theory from a symbolic interactionist framework with *role* as a negotiated set of relations. For students unfamiliar with this theoretical approach, this chapter provides a useful overview to key concepts used throughout the book. The next four chapters focus on women's familial roles as wives, mothers, kin members (daughters, sisters, and grandmothers), and homemakers. The fifth and sixth chapters explore a variety of roles outside home and family, including employer, employee, student, friend, neighbor, volunteer, and activist. The final chapter focuses on conflicts and strains that women face today.

One limitation of the book is Lopata's conceptual emphasis on women's familial and domestic roles. By devoting so much attention to these role clusters, Lopata appears to uncritically accept the notion that women's roles are defined by their relationships to their husbands or to their children. As a consequence, the book contains little or no discussion of heterosexual single women, lesbians, or women

who marry but decide not to have children. Despite this limitation, Lopata does an excellent job highlighting the conflicts experienced by a diverse group of married women with children through their involvement in modern and transitional roles. When both spouses work in highly demanding jobs, most wives and husbands fail to meet a modern arrangement with equal treatment of the demands of two jobs. Women still find themselves making adjustments to their husbands' careers and doing more child care and housework than their spouses do. According to Lopata, part of the problem is the lack of institutionalized support for women's modern roles. Though the burden of child care and housework still falls on women who work in the paid labor force, most workplaces are unwilling to accommodate women's familial responsibilities whether through institutionalized child-care arrangements, flexible work schedules, or family-related policies. While these conclusions are certainly not new, Lopata's book provides a terrific overview of the existing literature on women's varied roles in U.S. society and how these roles shift and change over the life cycle. As such, this is a fine book, a valuable resource for graduate students and a useful text for undergraduate courses on the sociology of family, gender, and the life course.

UNITED STATES POSTAL SERVICE™

Statement of Ownership, Management, and Circulation

(Required by 39 U.S.C. 3685)

1. Publication Title	2. Publication No.	3. Filing Date
SOCIAL FORCES	0 0 3 7 - 7 7 3 2	9-20-95
4. Issue Frequency	**5. No. of Issues Published Annually**	**6. Annual Subscription Price**
QUARTERLY (March, June, Sept., Dec.)	4	30.00 Ind. 52.00 Inst.

7. Complete Mailing Address of Known Office of Publication *(Street, City, County, State, and ZIP+4) (Not Printer)*

168 Hamilton Hall, UNC-CH, Chapel Hill, Orange County, NC 27599-3210

8. Complete Mailing Address of Headquarters or General Business Office of Publisher *(Not Printer)*

University of North Carolina Press, Box 2288, Chapel Hill, NC 27515-2288

9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor *(Do Not Leave Blank)*

Publisher *(Name and Complete Mailing Address)*

University of North Carolina Press, Box 2288, Chapel Hill, NC 27515-2288

Editor *(Name and Complete Mailing Address)*

Richard L. Simpson, 168 Hamilton Hall, UNC-CH, Chapel Hill, NC 27599-3210

Managing Editor *(Name and Complete Mailing Address)*

Paul Mihas, 05 Manning Hall, UNC-CH, Chapel Hill, NC 27599-3355

10. Owner *(If owned by a corporation, its name and address must be stated and also immediately thereafter the names and addresses of stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address as well as that of each individual must be given. If the publication is published by a nonprofit organization, its name and address must be stated.) (Do Not Leave Blank.)*

Full Name	Complete Mailing Address
University of North Carolina at Chapel Hill	Chapel Hill, NC 27599

11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check here. ☒ None

Full Name	Complete Mailing Address

12. For completion by nonprofit organizations authorized to mail at special rates. The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes: *(Check one)*

☒ Has Not Changed During Preceding 12 Months

☐ Has Changed During Preceding 12 Months
(If changed, publisher must submit explanation of change with this statement)

PS Form 3526, October 1994

(See Instructions on Reverse)

13. Publication Name	14. Issue Date for Circulation Data Below	
SOCIAL FORCES	9-20-95	
15. Extent and Nature of Circulation	Average No. Copies Each Issue During Preceding 12 Months	Actual No. Copies of Single Issue Published Nearest to Filing Date
a. Total No. Copies *(Net Press Run)*	4,433	4,300
b. Paid and/or Requested Circulation (1) Sales Through Dealers and Carriers, Street Vendors, and Counter Sales *(Not Mailed)*		
(2) Paid or Requested Mail Subscriptions *(Include Advertisers' Proof Copies/Exchange Copies)*	3,736	4,017
c. Total Paid and/or Requested Circulation *(Sum of 15b(1) and 15b(2))*	3,736	4,017
d. Free Distribution by Mail *(Samples, Complimentary, and Other Free)*	124	131
e. Free Distribution Outside the Mail *(Carriers or Other Means)*		
f. Total Free Distribution *(Sum of 15d and 15e)*	124	131
g. Total Distribution *(Sum of 15c and 15f)*	3,860	4,148
h. Copies Not Distributed (1) Office Use, Leftovers, Spoiled	573	152
(2) Return from News Agents		
i. Total *(Sum of 15g, 15h(1), and 15h(2))*	4,433	4,300
Percent Paid and/or Requested Circulation *(15c / 15g x 100)*	97%	97%

16. This Statement of Ownership will be printed in the December 1995 issue of this publication. ☐ Check box if not required to publish.

17. Signature and Title of Editor, Publisher, Business Manager, or Owner — Date

Judith Bergman

Judith Bergman, Assist. Director & Controller — 9-20-95

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions *(including fines and imprisonment)* and/or civil sanctions *(including multiple damages and civil penalties).*

Instructions to Publishers

1. Complete and file one copy of this form with your postmaster on or before October 1, annually. Keep a copy of the completed form for your records.
2. Include in items 10 and 11, in cases where the stockholder or security holder is a trustee, the name of the person or corporation for whom the trustee is acting. Also include the names and addresses of individuals who are stockholders who own or hold 1 percent or more of the total amount of bonds, mortgages, or other securities of the publishing corporation. In item 11, if none, check box. Use blank sheets if more space is required.
3. Be sure to furnish all information called for in item 15, regarding circulation. Free circulation must be shown in items 15d, e, and f.
4. If the publication had second-class authorization as a general or requester publication, this Statement of Ownership, Management, and Circulation must be published; it must be printed in any issue in October or the first printed issue after October, if the publication is not published during October.
5. In item 16, indicate date of the issue in which this Statement of Ownership will be printed.
6. Item 17 must be signed.

Failure to file or publish a statement of ownership may lead to suspension of second-class authorization.

SOCIAL FORCES

University of North Carolina Press
P.O. Box 2288
Chapel Hill, NC 27515-2288

Second class postage paid
at Chapel Hill, N.C.
and additional post offices

RETURN POSTAGE GUARANTEED

SOCIAL FORCES

AN INTERNATIONAL JOURNAL OF SOCIAL RESEARCH
ASSOCIATED WITH THE SOUTHERN SOCIOLOGICAL SOCIETY

VOLUME 74: NUMBER 3 MARCH 1996

IN GENERAL

1. Mail manuscripts (5 copies) to: **Social Forces**, Department of Sociology, Hamilton Hall, University of North Carolina, Chapel Hill, NC 27599-3210.

2. The paper must not be submitted to another journal while under review by SF.

3. Enclose a $15 processing fee. Papers are not reviewed until this fee is paid. The fee is waived for papers solely by currently enrolled students.

4. Manuscripts are not returned to authors.

5. Authors transfer copyright to SF. Permission fees for reprint of articles are split between authors and journal.

PREPARATION OF MANUSCRIPTS

SF Form For review, papers may be submitted in the form used by SF, American Sociological Association or other major sociological journals. Authors of accepted manuscripts are responsible for putting them into SF form.

Format, Organization, and Style Submit 5 clearly legible copies on 8 ½ by 11 good quality white paper. Double space typing throughout. For an accepted paper being revised to fit SF style, leave at least 1" margins; do not justify margins if using a word processor.

Order Organize the manuscript in this order: cover page; abstract; text; endnotes; references; tables; figures.

Cover Page Give title; author(s); affiliation(s); and a footnote (*) indicating name and address of the author to whom requests for offprints or other correspondence should be sent ("Direct correspondence to ___") and acknowledgment (if any) of financial or other assistance.

Abstract On a separate page, preceding the text, write a summary, 125 or fewer words (70 or fewer for a Research Note).

Endnotes Use only for substantive comments, bearing on content. Number consecutively from 1, double space, and append on a separate page.

Tables Type each table on a separate page. Insert a location note — "Table 1 about here" — in the text. For substantive footnotes relating to a table or figure use lowercase letters. Use tabs to separate columns.

Figures Must be artist-drawn, camera ready. Do not send original copy with a manuscript submitted for review. For review, figures may be roughly drawn but they must be neat and legible. For footnotes to figures, use same symbols as for footnotes to tables.

Mathematical Symbols (and others likely to be obscure to compositor) Clarify with encircled words written in the margin.

References in Text Indicate sources as illustrated below:

- when author's name is in text — Lipset (1960); when author's name is not in text (Lipset 1960)
- use page numbers only for direct quotations or specific notes or table — (Braudel 1969:213)
- for more than 3 authors use "et al."
- with more than 1 reference to an author in the same year, distinguish them by the use of letters (a,b,c) with year of publication (1975a)
- earlier publication should precede later publication in brackets with parentheses (Tocqueville [1835] 1956)
- enclose a series of references — in alphabetical order — in parentheses, separated by semicolons (e.g., Adler 1975; Adler & Simon 1979; Anderson, Chiricos & Waldo 1977; Bernstein et al. 1977; Chesney-Ling 1973a, 1973b).

References Following Endnotes List authors alphabetically, by surname. Spell out first names of all authors and editors. For authors with more than one work cited, list works earliest to latest. For articles, next give title of article (caps and lower case), name of journal, volume number, and pagination. For books and monographs, give title, followed by publisher.

Format of References Please spell out the first names of all authors and editors, unless they use only their initials or a first initial and a middle name in the source cited (e.g., Paul Radin, T.S. Eliot, and J. Owen Dorsey).

Elder, Glen H. 1975. "Age Differentiation and the Life Course." Pp. 165-90 in *Annual Review of Sociology*. Vol. 1, edited by Alex Inkeles, James Coleman, and Neil Smelser. Annual Reviews.

Myrdal, Gunnar. [1944] 1962. *An American Dilemma*. Harper & Row.

Ritzer, George. 1975a. *Sociology: A Multiple Paradigm Science*. Allyn & Bacon.

_____. 1975b. "Sociology: A Multiple Paradigm Science." *American Sociologist* 10:156-67.

Computer Files Authors analyzing data available in this format should cite their source, indicating producers and distributors as in these examples:

Louis Harris and Associates. 1975. Harris 1975 Nuclear Power Survey #2515 [computer file]. New York: Louis Harris and Associates [producer]. Chapel Hill: Louis Harris Data Center, University of North Carolina [distributor].

U.S. Bureau of the Census. 1979. 1970 Census of Population and Housing, Fourth Count Population Summary Tape [computer file]. Washington: U.S. Bureau of the Census [producer]. Rosslyn, Va.: DUALabs [distributor].

SOCIAL FORCES

MARCH 1996, VOLUME 74, NUMBER 3

ISSN 0037-7732 Library of Congress Catalog Number 24-31023

Second class postage paid at Chapel Hill, N.C. 27514 and additional mailing offices. *Postmaster: Send 3579 to Social Forces, UNC Press, P.O. Box 2288, Chapel Hill, NC 27515*

Claims for undelivered copies must be made within the month following the regular month of publication. The publisher will supply missing copies when losses have been sustained in transit and when the reserve stock will permit.

Social Forces is issued quarterly: September, December, March, June.

Social *Forces* is published by the University of North Carolina Press for the Department of Sociology, with the assistance of the Institute for Research in Social Science, University of North Carolina at Chapel Hill.

Rev. 2/2/96

INFORMATION FOR *SF* SUBSCRIBERS — For Volume 74

Social Forces is issued quarterly: September, December, March, June. Subscriptions are accepted on a four-issue basis only. We recommend that readers subscribe for a volume year — beginning in September and ending with the June number. Subscriptions entered after the appearance of the first issue in a volume should specify whether the subscriber wants the current volume (including back numbers) or wants to begin the subscription with the next volume.

Subscription Rates

Individuals - U.S.

Regular:	
1 year:	$30
3 Years:	$85
ASA:	
1 year:	$25
3 years:	$70
Students:	
1 year:	$12

Institutions - U.S.

1 year:	$52
3 years:	$145

Single Copies - U.S.
US $13

Individuals - Outside U.S.

Regular:	
1 year:	US $38
3 Years:	US $110
BSA, CSAA, ASA (non-U.S.):	
1 year:	US $33
3 years:	US $95
Students:	
1 year:	US $20

Institutions - Outside U.S.

1 year:	US $60
3 years:	US $170

***Social Forces* Cumulative Indexes**
Volumes 1-50 (1922-1972): **US $12**
Volumes 51-58 (1972-1980): **US $12**

(For single copies and the Index add US $1.50 for mailing outside the U.S.)

Back Numbers. Reprinted Volumes 1 through 40 are available from the Johnson Reprint Corporation, 111 Fifth Avenue, New York, New York 10003. Back numbers of Volume 41 and subsequent issues are available from the managing editor of *Social Forces*, Institute for Research in Social Science, Manning Hall CB #3355, University of North Carolina, Chapel Hill, North Carolina 27599. The four most recent back volumes are available from The University of North Carolina Press. Volumes are available in microfilm from University Microfilms, 300 North Zeeb Road, Ann Arbor, Michigan 48106.

Abstracts and Indexes. The journal's articles are indexed or abstracted in *A.B.C. Political Science & Government, Abstracts in Anthropology, Academic Index, Adolescent Mental Health Abstracts, American Bibliography of Slavic & East European Studies, Applied Social Sciences Index & Abstracts, Automatic Subject Citation Alert, Chicano Periodical Index, Communication Abstracts, Criminal Justice Abstracts, Current Contents, Current Index to Journals in Education, Current Literature in Family Planning, Education Administration Abstracts, Excerpta Indonesica, Family Studies Abstracts, Geographical Abstracts, Historical Abstracts, Leisure, Recreation & Tourism Abstracts, Linguistics & Language Behaviour Abstracts, Middle East: Abstracts & Index, PAIS: Public Affairs Information Service, Population Index, Psychological Abstracts, Social Work Research & Abstracts, Social Sciences Index, Social Sciences Abstracts, Social Science Citation Index, Sociological Abstracts, Studies on Women Abstracts, Urban Affairs Abstracts, Urban Studies Abstracts, and World Agricultural Economics & Rural Sociology Abstracts, International Bibliography of Periodical Literature*, and book reviews in *Book Review Index.*

Change of Address. Notice of change of address should reach the publisher at least 6 weeks before publication of the next number. Please include both old and new addresses and zip codes. *Communications regarding subscriptions should be addressed to: Social Forces* Subscriptions, The University of North Carolina Press, Box 2288, Chapel Hill, North Carolina 27515.
Editorial and business communications should be addressed to: Social Forces, 168 Hamilton Hall, University of North Carolina, Chapel Hill, North Carolina 27599-3210.

World Wide Web Page. Internet users can access information regarding *Social Forces*, including advertising rate card, authors' guide, and subscription form, at: http://www.unc.edu/depts/irss.sfhome.htm.

SOCIAL FORCES

MARCH 1996, VOLUME 74, NUMBER 3

Contents

Social Forces 74:3, March 1996

BOOK REVIEWS

Stratification Dynamics under State Socialism: The Case of Urban China, 1949-1993*

XUEGUANG ZHOU, *Duke University*
NANCY BRANDON TUMA, *Stanford University*
PHYLLIS MOEN, *Cornell University*

Abstract

Previous studies of stratification under state socialism have emphasized the institutional structure of redistributive economies. This article stresses the importance of the state and political mechanisms. We argue that individual life chances under state socialism are extremely sensitive to political processes, especially shifts in state policies, and that changes in state policies can dramatically alter opportunity structures, the status of structural locations, and the nature and value of political and human capital. Therefore, effects of these characteristics on life chances often vary across historical periods and over an individual's life course. Supporting evidence comes from a study of labor force entry and Communist Party membership in urban China from 1949 to 1993 based on life histories for samples of residents in Beijing and Shanghai.

Studies of stratification in industrialized market societies typically focus on inter- and intragenerational social mobility and status attainment (e.g., Blau & Duncan 1967; Breiger 1981; Goldthorpe 1987; Hauser & Featherman 1977). There has been extensive study of the channels of social mobility — the relative contribution to upward mobility of social status inherited from parents versus acquired human and social capital (Hout 1988; Yamaguchi 1983).

The focus on structural conditions, social origins, and occupational position implicitly assumes that the mechanisms of stratification are relatively stable

* *We are grateful to Liren Hou and Lin Lu for their able research assistance. The Sociology Departments at Fudan University in Shanghai and the People's University in Beijing, and, in particular Fan Weida, Li Qiang, Peng Xizhe, and Yan Jialin provided generous assistance with the data collection. We appreciate the helpful suggestions of David Grusky and Andrew Walder. This project has been partly supported by grants from the National Science Foundation (SES-9212936, SBR-9321462, SBR-9413540, SBR9541309), an ASA/NSF grant for the advancement of the discipline, the Halpern Fund, and an L.T. Lam Award from the East Asian Program at Cornell University. The first author was supported by a Spencer Fellowship from the National Academy of Education while working on this research. Send communications to Xueguang Zhou, Department of Sociology, Duke University, Durham, NC 27708, or via e-mail to xzhou@soc.duke.edu.*

across generations and over an individual's life course. In industrialized market societies, patterns of social mobility are relatively consistent over time and place (e.g., Erikson & Goldthorpe 1992; Grusky & Hauser 1984; Hauser & Grusky 1988; Sorensen 1992), even though some cross-national variation in mobility regimes clearly exists.[1] Father's social status and son's education, for instance, have remarkably consistent positive effects on intergenerational mobility over time (see, e.g., Featherman & Hauser 1978; Hout 1988).

Most of the existing stratification literature focuses on industrialized market societies, especially the U.S. While it provides a useful baseline, it has serious deficiencies for understanding stratification in state socialist societies. In these societies, stratification is organized around the state-socialist redistributive economy rather than a market economy. Resource transfers and the location of individuals in the social structure are largely determined by political power rather than by market mechanisms.

In this study we explore the linkage between the state and life chances in urban areas of the People's Republic of China. China is an excellent forum for such a study, not only because of its unique path of economic reform in the last 15 years, but also because of the marked political changes in its 45 years of communist rule. The Chinese experience is invaluable for studying the impacts of a highly volatile political and economic environment on individual life chances. By life chances, we mean events related to an individual's socioeconomic well-being, such as educational opportunities, position in occupational and organizational hierarchies, and political status.

To examine the links between the political dynamics induced by state policy shifts and individuals' lives in urban China, we analyze patterns of labor force entry and the acquisition of an important political status, Communist Party membership. We consider the impacts of not only the common determinants of life chances in industrialized market societies (e.g., social origins, inherited and achieved social and human capital) but also some distinctive factors in state socialist systems, in particular, sudden and sharp shifts in state policy. We use an event history framework to model changes in life chances over time.

Institutional Structure and Stratification

The idea that different political and economic institutions produce different patterns of stratification and inequality is not new. Ivan Szelényi (1978) argued that mechanisms of stratification in state socialist redistributive systems and in market systems differ qualitatively. He showed that in the state socialist system, the redistribution of resources and the allocation of welfare are based on political power, which engenders conflict between "immediate producers" and "redistributors." The linkage between the state and individual life chances has been widely recognized by scholars studying China (Blau & Ruan 1990; Lin & Xie 1988; Parish 1984; Whyte 1985). Titma and Tuma (1993) also called attention to the political construction of opportunity structures in state socialist societies, stressing the importance of the hierarchies of economic sectors (ministries) and of places in the stratification system in the Soviet Union. The central role of the state and of political power as stratifying mechanisms in state socialist societies

leads their labor market structures to differ markedly from those in market societies (Burawoy & Lukács 1985; Connor 1979; Stark 1986).

Consider the socioeconomic status of occupations in Ivan Szelényi's (1978) study of Hungary. Although the occupational categories are similar to those in industrialized market societies (i.e., professionals, managers, clerks, skilled and unskilled workers), the stratification mechanisms differ substantially. The privileges and benefits of various social groups depend systematically on their relationship to the communist state. For instance, the communist state bred a new "bureaucratic class" of redistributors who enjoyed greater political power and a higher socioeconomic status than all other social groups (Bauman 1974; Djilas 1966; Konrád & Szelényi 1979). Although Szelényi discussed Eastern Europe, his argument fits China well (Nee 1989). The dominance of the redistributive system in China parallels that in Eastern Europe, and its bureaucratic class enjoys special privileges and benefits similar to its East European counterparts (Zhou 1995). Walder (1995) also finds that professionals and bureaucratic managers have distinctive career paths.

Recently Lin and Bian (1991) and Walder (1986, 1992) have called attention to another structural feature of stratification under state socialism, the differentiation of organizations based on state socialist property rights. In China, the relations of organizations to the state are hierarchically ordered in terms of property rights, which causes these organizations to differ in political power, economic benefits, and prestige. The organizational hierarchy is a key feature that helps legitimate the state socialist system. As Walder (1986) says: "The extraordinary job security and benefits, the goods and services distributed directly by the state enterprise in a situation of scarcity that affects other sectors of the workforce more severely, is an important source of acceptance of the system" (248).

Since the distribution of resources and welfare occurs mainly through organizational channels, individuals in different types of organizations have different career trajectories, promotion opportunities, and other life chances.[2] Those working in government agencies, for instance, have not only authority over those in other types of organizations but also better access to housing, hospitals, and so on. Lin and Bian (1991) found that urban China has a segmented labor structure based on the organizational hierarchy and that this hierarchy significantly affects status attainment. Even for workers, social status depends on the property rights relation of their organization to the state. Workers enjoy better welfare programs in state-owned enterprises than they would in the collective and private sectors (Bian 1994; Walder 1992).

The Dynamics of Social Stratification

Although the unique institutional structure in state socialist societies is important, an emphasis on structural conditions alone is incomplete for understanding stratification in these societies. In contrast to industrialized market societies, in China the structural conditions — occupational and organizational hierarchies — are subject to change, especially to political changes directed by the communist state and implemented in state policies.

Stratification under state socialism is constructed through a top-down process. Hence, the impact of political change on individual life chances depends on both the scope and timing of state policies.

The broad scope of the socialist state's power over all areas of social life distinguishes it from other powerful and repressive states that nonetheless have relatively small effects on individuals' everyday lives. From income distribution, to housing, to various welfare programs, almost all life chances of individuals and social groups in China are under the distributive power of the state (Walder 1986; Whyte & Parish 1984). Equally critical is the timing of state policy shifts. Given monopolistic power concentrated in a few leaders, the communist state can transfer resources from one sector to another relatively quickly, either in response to emerging crises or in pursuit of the leaders' goals. Consequently, the macropolitical process is often characterized by sharp political twists and turns in state policies.

To the best of our knowledge, Davis (1992a, 1992b) conducted the first elaborate and systematic study in urban China of the inconsistent paths of individuals' lives resulting from political changes. She demonstrated that downward mobility of the children of the urban "middle class" during the Cultural Revolution was largely due to shifts in state policies. Similar patterns have been observed in other state socialist societies. For example, Szelényi and Manchin (1987) noted that the state's repressive policy of collectivization interrupted embourgeoisement in rural Hungary in the earlier period and called explicitly for a "life history" perspective to study it (see also Szelényi 1988). However, the structural conditions of state socialism have overshadowed the dynamic aspects in most of the existing literature.

Social positions are associated with different political, ideological, and economic resources. Stratification structures channel resource transfers across generations and over stages of an individual's life course. Weber (1978) views the stratification structure as ultimately resting on market position: "The kind of chance in the *market* is the decisive movement which presents a common condition for the individual's fate. Class situation is, in this sense, ultimately market situation" (928). Thus, to understand stratification from a comparative perspective, one needs to understand differences in market and redistributive mechanisms in transferring resources across generations and in allocating resources among social groups

In a market economy, political and economic transactions are based on the principle of *exchange* (Lindblom 1977). In this institutional structure, initial endowments of resources have a lasting effect on social position because private property rights prevail and stabilize the relative opportunities of various groups based on their preexisting social and economic resources. Not surprisingly, then, until very recently social mobility and stratification in industrialized market societies have been characterized by persistent advantages of initial resource endowments. Studies invariably find that social mobility in industrialized market societies mainly operates through individual and family-based social inheritance and achievement. Improvements in socioeconomic status have largely resulted from the emergence of new opportunities due to industrialization and concomitant structural changes (Treiman 1970). The implication for

individual life chances in market systems is a high degree of stability of the relative social status of social groups.

Mechanisms of resource transfer in state socialist societies are qualitatively different. Under state socialism, boundaries and the relative positions of social groups (classes), organizations, and localities are based on state policy considerations rather than market transactions (Zhou 1993). As a result, social boundaries are fluid and vulnerable to state policy shifts. In tandem with state policy shifts, resources are redistributed across sectors, and the life chances of social groups and individuals in various social sectors change. In sum, macropolitical dynamics decisively and directly affect life chances of social groups and individuals in China, regardless of their structural locations.

Specifically, we point to two distinctive effects of state dominance on patterns of stratification. First, the state may alter the relative socioeconomic status of social groups by drastically shifting their relative resources. Such resource transfers are typically based on ideological and political considerations rather than on market efficiency. During the 45 years of communist rule in China, state policies have changed frequently and substantially, for example, between political control and economic development, between competence and political loyalty in criteria for promotion, and between centralization and decentralization in administrative structure (Nathan 1976; Zhou 1992). Drastic shifts in the relative status of social groups have resulted. For example, state policies that redirected resources to children of the working class (e.g., educational opportunities and recruitment into the Communist Party and into higher-status organizations and jobs) substantially changed patterns of social mobility in China (Parish 1984). It should be noted, however, that the state policy of discriminating against certain classes was officially discontinued in 1979.

Second, in a politicized society, the symbolic boundary between "us" and "them" is critical, and shifts in state policies often redraw such boundaries. Indeed, state policies may change the very meaning of "resources," thus altering the nature of social inheritance. For example, education was a political liability rather than a resource during the Cultural Revolution, and white-collar occupations have not consistently delivered the economic benefits and prestige common in other societies (Davis 1992a). Thus, to understand resources associated with positions in the social structure requires substantive interpretation of the historical context.

Consequently, we expect that the distribution of opportunities in China varies systematically with shifts in state policies. State policy shifts often expand or reduce the opportunities in ways that affect all social groups and individuals simultaneously. In all societies, external shocks (e.g., economic crises, war, natural disasters) adversely affect all social groups, but the extent varies due to social differentiation of groups (Elder 1974). In China, drastic changes in overall opportunities have often been due to disastrous state policies. As a result, their effects have been greater, less predictable, and more uniform. Parish (1984) demonstrated the systematic effort of the Chinese state to pursue an egalitarian society during the Mao era. Similarly, in discussing the "politics of life chances" in China, Whyte (1985) traced four historical periods under communist rule and observed that state policies tend to affect all social groups in comparable ways. Sometimes they raise the general level of economic development and opportuni-

ties, as occurred during the first period, the 1950s, when expansion of the state and the economy created more opportunities and widespread upward mobility. Contrarily, policy disasters (e.g., the Great Leap Forward and its aftermath in 1958-62) reduce living standards and economic development and adversely affect the life chances of individuals across organizational and occupational boundaries.

In addition, we think that effects of individual attributes on life chances vary over time. The literature on stratification dynamics in industrialized market societies rests on the assumption that human, social, cultural, and political capitals (e.g., parental socioeconomic status, education, political affiliation) invariably have positive returns and enhance individuals' upward mobility. Chinese political dynamics imply that conventional capitals embedded in individuals do not have consistent impacts on life chances over time: In the Chinese experience, today's capital may become tomorrow's liability.

Consider education as an example. In industrialized market societies, education contributes significantly to individuals' upward mobility (Blau & Duncan 1967; Hout 1988). In China, however, the effect of education on individuals' life chances has varied dramatically depending on current state policies (Kraus 1981; Rosen 1982; Unger 1982). In the early years of communist China, the Communist Party stressed candidates' "revolutionary experience" (Lee 1991) and downplayed education. During the Cultural Revolution, Maoist leaders treated intellectuals with suspicion and hostility and made serious attempts to recruit those with peasant or worker backgrounds into management positions, viewing them as supportive of the radical political cause (Whyte 1973). In the post-Mao era, education has again become important for upward mobility. The new cadre policies adopted in 1980 emphasize educational qualifications and competence in recruitment and promotion (Lee 1991). In a fluctuating environment, human, social, and political capital do not necessarily accumulate and give consistently positive returns.

In a similar vein, we also think that the effects on life chances of structural position in the occupational and organizational hierarchies vary over time. As noted above, the occupational and organizational hierarchies in China are constructed by political processes. In contrast to the gradual expansion of opportunities with industrialization in market societies, shifts in state policies in China dramatically alter life chances of social groups from time to time.

For instance, the social status of occupational groups is politically constructed in accord with official class identification in Communist ideology. The high political status of industrial workers (*chanye gongren*) was not based on their prestige or income but on ideological considerations of the communist state. But as a political construction, the status of occupational groups is sensitive to state policy shifts (Xiao 1989). Workers were favored in Mao's era but their political status declined substantially during the reform era. In contrast, intellectuals had low status in Mao's era but a higher status in the reform era. Similarly, although the organizational hierarchy has been relatively stable over time, its effect on life chances has not been constant. During the Cultural Revolution, individuals in all kinds of organizations experienced similar dislocations. These dramatic changes across historical periods suggest nonmonotonic effects of structural conditions on individual life chances.

In sum, our discussion points to the crucial importance of the *dynamics of stratification* under state socialism. By this term we mean the patterns of change in stratification structures (e.g., in the location of boundaries, in relative rankings of social positions) and in the mechanisms of stratification. Our central argument can be summarized as follows: Life chances under state socialism, in contrast to industrialized market societies, are affected not only by the distinctive structural conditions of the state redistributive system (the occupational and organizational hierarchies), but to a much greater extent by political dynamics and concomitant shifts in state policies that directly and immediately alter the opportunity structure and the value of various social resources for individuals' life chances. Political dynamics are not irrelevant to life chances in industrialized market societies, but shifts in opportunity structures and mechanisms evolve much more slowly and smoothly in such societies, making it possible to focus on structure alone in most analyses of industrialized market societies.

Data

Our empirical analyses are based primarily on life history data for a sample of 1,038 urban residents in Beijing and Shanghai. The data were collected in the summer of 1993 by interviewers using a pretested questionnaire. The questionnaire collected life history data on the respondent's shifts across organizations, occupations, and geographic locations. We also asked about family background, family structure, and political status over time, as well as income and housing conditions in selected years.

Since we collected the data retrospectively, errors in recalling events are inevitable. An advantage of collecting event history data in China is that individual job mobility was severely limited before the economic reforms in 1979; hence respondents did not need to recall many changes. Mobility has increased considerably in the last decade, but recent events are likely to be recalled fairly accurately. All in all, errors from collecting the data retrospectively are likely to be less serious in China than in some other societies.

We used a multistage sampling scheme to draw our sample in the two cities. In China, a metropolitan city is composed of residential districts. Each district is composed of residential "streets," within which households are organized into residential blocks (*Juweihui*). In each city we selected first residential districts and then residential blocks based on their proportion of residents, according to official statistics provided by the municipal government. To select households within each residential block, we used a systematic random sampling procedure; that is, we chose every *n*th household. In all, we chose 1,300 households in a total of 75 residential blocks in six districts. Finally, a member of the household between ages 25 and 65 was chosen for the survey using a random number table. Respondents included both current labor force participants and retirees. We successfully completed 1,038 interviews, giving a response rate of 80%. Interviewers rated almost all respondents as cooperative.

We note that the two cities are not representative of urban areas in China. Beijing, the political capital, and Shanghai, the largest industrial center in China, certainly have enjoyed better opportunities than other urban areas. Moreover,

in a highly centralized political system like China's, state policies can produce marked variations among both regions and urban areas. Therefore, our analyses provide only a glimpse of the dynamics of stratification in urban China. Nevertheless, given the strong grip of the socialist state, changes in life chances across urban areas in China are likely to be much smaller than those in a market economy. Therefore, we believe the results of our analyses are indicative of the overall patterns in urban China.[3]

Variables

Our discussion of stratification dynamics under state socialism points to varying mechanisms of stratification, on the one hand, and varied opportunities contingent on state policy shifts over time, on the other. Along these lines, we consider three dependent variables (first entry into an occupation, first entry into a type of organization, entry into the Communist Party), three main independent variables (human capital, political capital, historical periods), and three control variables (gender, place, age).

DEPENDENT VARIABLES

Occupation

We used the following occupational groups, ordered according to their hierarchical position in the state socialist redistributive system: (1) Cadres include those in managerial positions in enterprises as well as cadres in party and government agencies. (2) Professional jobs are occupations similar to those in industrialized societies, such as university faculty, teachers, and researchers. (3) Workers include all types of workers (including those in services and transportation), regardless of their position in the organizational hierarchy. (4) Farmers (peasants) in these two cities are comprised mainly of the "sent-down" youth who later returned to their original urban areas. Due to the rarity in our sample, we treated entry into two other occupations, self-employed entrepreneurs and clerical workers, as censored observations in our analyses of entry into an occupation, thereby avoiding the biases that would result from simply dropping them from the analysis.

Type of Organization

Our measures of organizational type are based on two dimensions: social-political status and the vertical property-rights relation to the state. We distinguish the following types of organizations,[4] listed roughly in descending hierarchical order:

(1) Government agencies (*zhengfu Jiguan*) include ministries, commissions, bureaus, and offices at various levels of the Communist Party and of state bureaucracies. Both the Communist Party and the administrative apparatus, which are interwoven at every level, belong to this category. Government agencies participate in making and/or implementation of state policies.

(2) Public organizations (*shiye danwei*) are nonprofit organizations in the public domain. They include educational and research institutions, and

organizations in the medical, publishing, broadcasting, and entertainment sectors. Although not administrative organs of the state, most of these organizations are closely linked with the state or with local governments in two respects. First, these organizations are financed directly by the governmental budget; secondly, the personnel systems in these organizations are under the same administration of the central government as those in government agencies. For instance, educational institutions are financed by the central or local government and their personnel system is directly subject to the regulation of the corresponding government agencies. Until recently, under the central planning system, those employed in government agencies and those working in public organizations were covered in the same salary system. That is, employees in both types of organizations had the same opportunities for salary increases. Certain types of public organizations, such as elementary and junior high schools, may not be as favorable in terms of welfare as some state-owned firms. Overall, however, public organizations as defined here enjoy higher socioeconomic status under the state redistributive system.

(3) State-owned firms (*guoying qiye*) include work units in manufacturing, processing, and other production areas, as well as ones in service sectors owned by central or municipal governments. These organizations enjoy extensive welfare programs in housing, health care, and welfare subsidies.

(4) Collective firms (*jiti qiye*) are firms not directly under the administration or financial support of central planners, though they are often sponsored by a local government (e.g., a district or county government or residential office). This type of organization offers the least desirable welfare programs and had low social status in urban China until very recently.

(5) Villages or people's communes in rural areas are included as a type of organization because they were the main destination of "sent-down" youth during the Cultural Revolution.

Party Membership

Respondents were asked if they had ever joined the Communist Party and, if so, in what month and year they had joined.

INDEPENDENT VARIABLES

Human Capital

We follow convention and measure human capital by educational attainment. Educational level is classified in terms of: (1) college (including *dazhuan* and graduate programs); (2) high school (including technical high school, *zhongzhuan*); (3) junior high school; and (4) elementary school or below. We collected information on both the timing and level of educational attainment.

Political Capital

In China, "family background" *(jiating chushen)* refers to the Communist Party's grouping of individuals into different classes according to their family roots. Although highly correlated with father's socioeconomic status, "family background" is not the same as father's occupation currently or at age 14. Rather, it

refers primarily to the *class* position of the father or grandfather before 1949.[5] We use "family background" instead of the conventional father's social status because this politically defined category has been an important criterion in job assignments and promotions.

Based on our knowledge of this political construction, we classified family background into the following four categories of descending status: (1) The *cadre class* refers to those whose parents are cadres. This group has been argued to enjoy special political privileges under state socialism. (2) The *working class* includes adult children or grandchildren of workers or poor peasants; it is the group that the Communist Party proclaimed to be the cornerstone of communism. (3) The *"distrusted" class* includes those parents or grandparents who were small business owners (*yezliu*) or middle peasants (*zhongnong*), as well as those whose parents or grandparents were landlords or large business owners (*zibenjia*).

Historical Periods

Our discussion of the processes of stratification in China implies strong period effects, which are highly associated with state policies and which indicate the overall level of opportunities for all social groups (Whyte 1985). We identify three broad historical periods: 1949-65, 1966-79, and 1980-93[6] and discuss the historical context in these periods in the next section.

Control Variables

Control variables include gender (a dummy for being female), place (a dummy for Beijing rather than Shanghai), and age (measured in years minus twenty).

Given the nature and scope of our theoretical concerns, we analyze data only on individuals who entered the labor force after 1949. Cases with information missing on any covariate in a particular model are omitted from that specific analysis. The number of cases dropped because of missing data is very small. In analyses of entry into an occupation and organization, we included observations only of those older than 13, assuming that labor force entry at younger ages might signal a respondent error. In the analyses of entry into the Communist Party, we excluded observations after the individual had retired from the labor force and before all individuals reached age 17 (because 18 is the minimum age formally allowed for Communist Party membership). We also excluded observations on individuals in the farm and private sectors because these two groups experienced few events and are substantively different from the reference category of workers.

Models and Methods

We use event history analysis (Tuma & Hannan 1984) to examine individual life chances, life-course trajectories, and rates of transition among differentially ranked social positions. An important advantage of event history analysis is that information on previous life history and on time-varying covariates can be incorporated in analyses of the time-dependent rate of occurrence of various life events.

The general modeling strategy is as follows. Suppose there are data on a sequence of life events over time, such as enrollment in educational institutions, assignment to a particular type of occupation and organization, promotion, and so on. Observed events are regarded as realizations of a latent, random process of change in individual life chances that depends on covariates of theoretical interest, x(t), which may change over time. Following Tuma and Hannan (1984), we define instantaneous transition rates as follows:

$$r_{jk}(t \mid w_{n-1}) = \lim_{\Delta t \downarrow 0} \frac{q_{jk}(t, t + \Delta t \mid w_{n-1})}{\Delta t}$$

where $r_{jk}(t)$ is the instantaneous rate of transition from status J to status k at time t, $q_{jk}(t, t + \Delta t)$ is the probability of a transition from j to k between time t and $t + \Delta t$, and w_{n-1} is the previous history. Since we examine the first event, the number of the event, n, is 1 in our analyses, and previous history refers to the individual's family background.

Our argument about inconsistent effects of human and political capital over time implies a need for models that allow the effects of the covariates of theoretical interest to vary across historical periods. For this purpose, we use piecewise exponential models that allow intercepts and effects of covariates to vary across historical periods:

$$\log r_{jkp}(t) = \beta'_{jkp}\, x(t)$$

where the subscript p denotes a given historical period (e.g., 1949-65, 1966-79, or 1980-93), and $x(t)$ refers to the set of explanatory and control variables used in the analysis. We note that the values of some variables (in particular, age and educational level) change over time for a given individual. In the analysis of entry into the Communist Party, employment status (organization and occupation) also changes over time.

We estimated models by the method of maximum likelihood, which lets us test nested models using likelihood-ratio χ^2 statistics (Tuma & Hannan 1984). The standard errors of parameters in a correctly specified model are asymptotically normally distributed, which lets us test hypotheses that any given parameter in the model differs from some prespecified value (usually zero).

In some analyses we constrain the coefficients of x(t), except for the intercepts, to be the same in every historical period; we denote this as "model A." In other analyses, we let all parameters vary across the historical periods; we call this "model B." Since model A is nested within model B, we can compute a likelihood ratio χ^2 statistic to test whether the effects of covariates vary significantly across the historical periods.

Results

We argued above that shifts in state policies can affect individual life chances in several conceptually distinct ways. First, state policy shifts can directly alter the opportunity structure — the distribution of various resources, social positions, and rewards. The state can expand or reduce access to certain educational levels, occupations, organizational types, and political positions, such as

membership in the Communist Party. In addition, the state can change the criteria used in allocating people to various social positions, as well as alter the political and social status of different social groups. This consequence of state policy changes is the main concern of our research. Finally, via the previous two effects, state policy changes can indirectly induce effects on the composition of the labor force, and thereby amplify or dampen the impacts of other changes. In trying to understand the relationship between stratification and political dynamics, one needs to take all sources of change into account. Below we summarize what we have learned about each source of change based on available public documents and our survey data. Given our emphasis on the fluctuating political environment and state policies, we begin by introducing the historical context of state policy shifts and then describe the evolution of opportunity structures in urban China.

STATE POLICIES AND THE HISTORICAL CONTEXT

Since the establishment of the Communist government in China in 1949, Chinese political and economic structure has been based largely on a Soviet-style party-state. China experienced increasing expansion of the central government and the elaboration of socialist redistributive system (Whyte & Parish 1984). In the 45 years of Communist rule, however, China also experienced frequent and large-scale political and economic changes induced by drastic shifts of state policies. The Anti-Rightist Campaign of 1957, the Great Leap Forward in the late 1950s, the Cultural Revolution, and the recent economic reform are well-known examples. These large-scale socioeconomic changes have had profound impacts on the redistribution of opportunity among social groups. To explore the effects of political dynamics on life chances, we distinguish three broad historical periods that characterize drastic shifts in state policies.

The period 1949-65 was one of nation building and economic development. The Communist government centralized resources and its power through collectivization and nationalization, especially in urban China (Whyte & Parish 1984). During this period, state labor policies were largely geared toward industrialization and economic development, leading to overall expansion of opportunities, but there were notable political fluctuations and political campaigns (Whyte 1985).

The period 1966-79 was the period of the Cultural Revolution and its immediate aftermath. There were not only large-scale political purges and repression, but labor force participation and other opportunities that were highly politicized and severely affected by sharp shifts in state policies. In particular, political standards became the main criterion for job allocation, promotion, and the distribution of resources among social groups. During most of this period, regular promotions and wage increases were frozen.

The period 1980-93 is the post-Mao era when most policies adopted during the Cultural Revolution have been reversed and various reforms have been introduced, especially in urban areas. The economic reforms in this period introduced market mechanisms in resource allocation and created new opportunities, leading to the erosion of the redistributive state (Nee 1989, 1991).

However, since the reforms did not begin in urban areas until 1985 and happened erratically, except in some coastal areas, the effects of an emerging market on labor force participation have been limited in urban China.

CHANGES IN THE OPPORTUNITY STRUCTURE

We have argued that individual life chances vary over time in response to altered opportunity structures resulting from shifts in state policies. The historical patterns of educational and labor force participation, of labor force entry, and of joining the Communist Party provide some evidence that supports our argument.

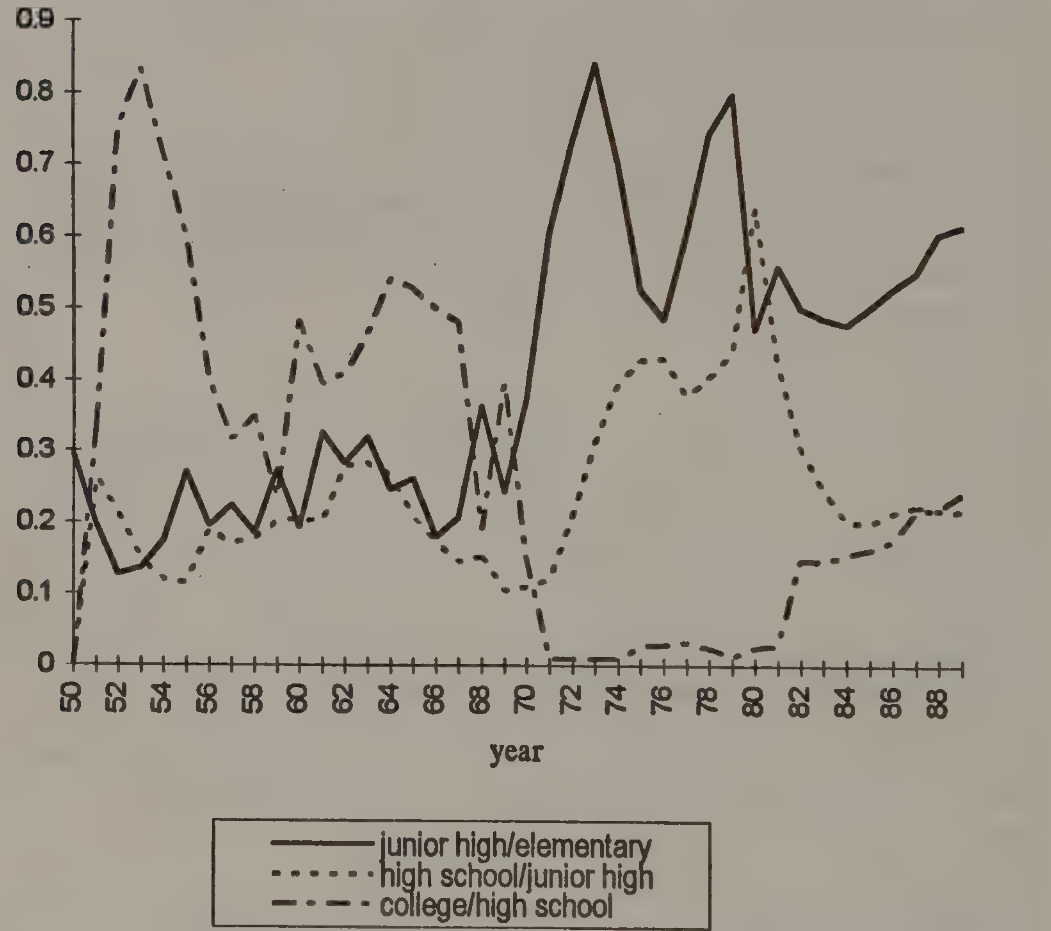
Educational Opportunities

Figure 1 shows the number of graduates at each educational level divided by the number at the level just below it by year for 1949-89.[7] It provides evidence that educational opportunities in China have varied across historical periods in accordance with state policies. First, the proportions of high school and college graduates were extremely low during the earlier years of the Cultural Revolution when radical state policies severely disrupted the educational sector. Second, the extremely low proportion of college graduates from 1971 to 1981 resulted directly from state policies that closed colleges during the early years of the Cultural Revolution. In fact, no high school graduates, regardless of their social origins, were directly admitted to college from 1966 to 1978. The first cohort of college students after the Cultural Revolution did not graduate until 1982.

Labor Force Participation

Figures 2 and 3 show the labor force composition from 1949 to 1989 in Beijing and Shanghai, respectively (China State Statistics Bureau 1990). (Private entrepreneurs are not displayed due to their small numbers.) A comparison of the two figures reveals a striking similarity between the two cities over time. Such similarity indicates the impact of state policies. For instance, the policy to incorporate private enterprises into state-owned ones after 1956 led to a sharp rise in employment in the state sector. Similarly, the economic crises of the early 1960s led the government to force large numbers of workers to return to rural areas; it is reflected in a sharp decrease in employment in the state sector in the early 1960s. In 1968, Mao Zedong and the Communist leadership instituted the "send-down" policy, which forced junior high and high school graduates (the "educated youth") to the countryside. Between 1968 and 1979, over twenty million educated youth were sent to rural areas to work as farmers. Roughly a half returned to urban areas before 1978, and the rest mainly returned by the end of 1979, when the original policy was denounced (Deng 1993). Note that the proportion of the labor force in state sectors compared to that in the collective sectors is relatively stable, even in the post-Mao era.

FIGURE 1: Ratio of Graduates from Each Educational Level (By Year)



Changes in Access to a First Occupation and First Organization

We next consider changes over time in work opportunities as estimated from the data on our sample. We consider access to a first occupation and a first organization in the three historical periods. One of the most revealing statistics for this purpose is the overall annual rates of entry into various occupations and organizations in the three historical periods, which are reported in Table 1.

The annual rates given in this table clearly indicate that opportunities have changed markedly across historical periods. First, there is evidence of long-term trends in industrialization and economic development: The rates at which labor market entrants become workers and professionals are much higher in the 1980-93 period than in 1949-65. Correspondingly, the rates of entering public organizations (where many professionals work) and state-owned firms are also markedly higher in 1980-93 than in 1949-65. The rate of labor market entrants entering a government agency also grows substantially, but more modestly. These shifts reflect the political leadership's renewed concern for economic expansion after the Cultural Revolution.

However, not all changes in work opportunities can be explained by economic development. For most occupations and organizations, changes over time show no clear trend. In terms of entry into an occupation and organization,

TABLE 1: Structure of Opportunities by Historical Period: Entry Rates per Year for First Occupation, First Organization, and the Communist Party

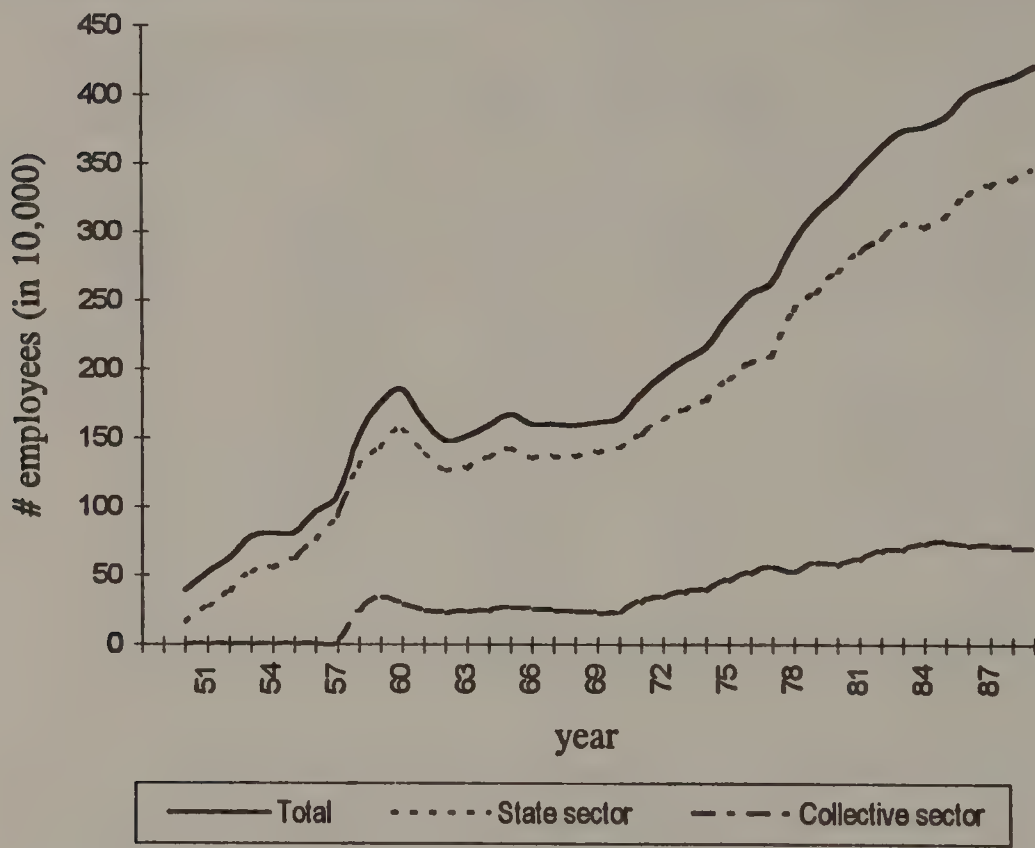
	All Years	1949-65	1966-79	1980-93
Occupation				
Cadre	.016	.018	.014	.016
Professional	.025	.024	.013	.060
Worker	.082	.066	.089	.106
Farmer	.018	.002	.044	—
Type of organization				
Government agency	.018	.019	.013	.024
Public organization	.025	.025	.014	.051
State firm	.068	.054	.068	.107
Collective firm	.015	.010	.020	.020
Village	.018	.002	.044	—
Communist party	.009	.013	.009	.007

the distinctiveness of the Cultural Revolution (1966-79) is striking. It is the only period with nonnegligible rates of labor market entrants becoming farmers and moving to a village type of organizations.[8] This period is also noteworthy for lower rates of labor market entrants obtaining a high-status job — becoming a professional or (to a lesser extent) cadre — and for moving to a government agency or public organization. These changing patterns across historical periods undoubtedly reflect the radical policies introduced during this period.

Changes in Access to the Communist Party

We next consider access to membership in the Communist Party, a very important political resource for people living in urban China. The findings based on our data, shown in the last row of Table 1, are somewhat surprising: Rates of entry into the Communist Party fall steadily across the three historical periods — the rate in the 1980-93 period is about one-half that in the 1949-65 period. One must be careful in interpreting this decline. A falling entry rate does not necessarily mean that the proportion of the population belonging to the Communist Party falls. The rate in the early period may have been unusually high simply because the Communist Party had relatively few members in 1949 and sought to enhance its legitimacy by expanding its membership rather quickly. After the initial rapid expansion, fewer new members were needed each year to maintain a steady size since membership in the Communist Party is usually for life. Still, the decrease means that later cohorts had less access to this important political resource than did earlier cohorts who were eligible to join the Party in the early years of the communist state.

FIGURE 2: Labor Force Composition in Beijing, 1949-1989



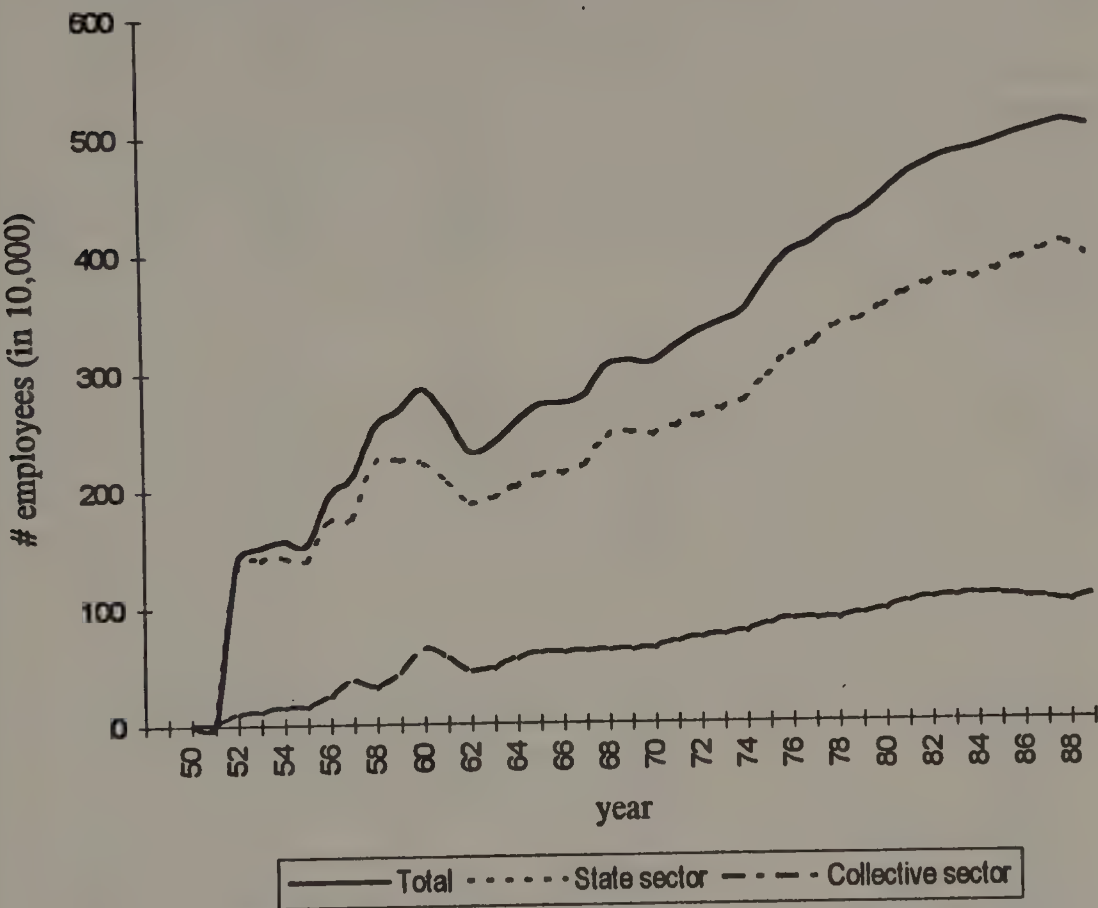
Summary

Overall, changes over these three historical periods in educational and labor force participation and in the entry rates reported in Table 1 reveal major variation in the opportunity structure in urban China during this 45-year period. The differences in opportunities in various occupations and types of organizations during the Cultural Revolution are so sharply different from any trend expected because of economic development that it is reasonable to attribute them to shifts in state policies. The conclusions are obvious: Changes in opportunity structures directly affect the distribution of life chances for individuals who entered the labor force in a particular historical period

Changes in Composition of the Population

Using data on our samples in Beijing and Shanghai, we next consider cross-period variation in key characteristics of the populations at risk of entering various positions in the main social hierarchies (see Table 2).[9] The variation in the population's composition are indicative of the consequences of the twists and turns in state policies in urban China.

FIGURE 3: Labor Force Composition in Shanghai, 1949-1989



Composition of Population at Risk of Entering the Labor Force

Clearly, the composition of people at risk of entering the labor force (shown in the top half of Table 2) changes strikingly over time. The period of the Cultural Revolution stands out. Among those at risk of entering the labor force, the proportion of college graduates and the average age are lower in 1966-79 than in the other two periods. These two compositional changes are not unrelated. The closing of colleges and universities and the policy to shorten the time children spent in elementary and junior high school during this period not only reduced the proportion of college attendees and college graduates but also pushed young people into the labor market at younger ages.

Other variations over time in the composition of those at risk of entering the labor force are more subtle consequences of earlier political changes. Following the establishment of communist rule and the concomitant elimination of certain occupations found in the old system and the creation of new occupations (especially cadre), the family background of those at risk of entering the labor force changes dramatically over time. After thirty years of communist rule — that is, since 1980 — most young labor market entrants had parents who acquired positions under the communist system. Consequently, by the 1980-93 period, the proportion with a distrusted family background has shrunk

TABLE 2: Means of Covariates for Risk Sets by Historical Period

	1949-65	1966-79	1980-93
Entry into the labor force			
Education			
Elementary	.40	.17	.06
Junior high	.33	.54	.28
Senior high	.19	.25	.40
College	.09	.04	.26
Family background			
Cadre	.05	.15	.39
"Distrusted"	.28	.15	.11
Worker	.67	.70	.50
Age	18.27	17.65	20.26
Female	.57	.48	.48
Beijing	.52	.43	.67
Number of individuals at risk	527	578	195
Entry into the communist party			
Education			
Elementary	.43	.23	.11
Junior High	.18	.36	.37
Senior High	.22	.26	.35
College	.17	.14	.18
Family background			
Cadre	.02	.06	.14
"Distrusted"	.29	.24	.18
Worker	.69	.69	.68
Occupation[a]			
Cadre	.07	.09	.07
Professional	.14	.14	.14
Soldier	.06	.02	.005
Workers and others	.73	.77	.79
Type of organization[a]			
Government agency	.16	.07	.05
Public organization	.16	.19	.17
Collective firm	.08	.12	.13
State firm and others	.60	.63	.64
Not yet entered labor force	.44	.15	.07
Age	24.26	30.01	35.06
Female	.57	.55	.51
Beijing	.56	.45	.44
Number of individuals at risk	367	695	831

[a] Distribution excludes those who have not yet entered the labor force.

enormously, while the proportion with a cadre family background has grown greatly. Note that the great growth in a cadre family background may be higher in our study cities of Beijing and Shanghai than in other Chinese cities because of the leading role of these cities in China's political and economic system.

The broad expansion over time in the educational level of labor market entrants, which is another result of state policies, is also readily apparent in Table 2. The proportion who have graduated from a senior high school rises steadily, while the proportion with elementary or no education falls markedly. Among those at risk of entering the labor force, the proportion who have graduated from junior high school rises and later falls, as the educational system continues to expand.

Composition of the Population Eligible to Join the Communist Party

Changes over time in the composition of the group at risk of entering the Communist Party are roughly similar to the changes for the group at risk of entering the labor force. However, the changes in the former group are more muted because the population at risk is much larger and much more diverse, especially in age.

Since our analyses of entry into the Communist Party include measures of labor force status (in urban areas), Table 2 reports compositional changes in it as well as education, family background, and our control variables. The two largest changes are in average age and in the proportion who have not yet entered the labor force. The former is largely a result of our sample design. Many people who were at risk of joining the Communist Party in 1949-65 were no longer living in 1993 or older than 65, the upper age limit for selecting the sample. Given the relative youth of this group in 1949-65, the very high proportion who have not yet entered the labor force during this period is less surprising. It may also be unusually high in 1949-65 because many women did not participate in the paid labor force period to 1949. Even with these two cautionary notes, the change over time in average age and in the proportion who have not yet entered the labor force is striking.

The proportion of those in the labor force who are in various occupations does not change much across the three periods, although the proportion who are soldiers falls markedly and the proportion who are workers or in other, peripheral occupations rises correspondingly. The proportion of those in the labor force who are in various types of organizations changes somewhat more, with the fraction working in a government agency falling and the fraction working in collective and state-owned firms rising slightly.

CHANGES IN ALLOCATION CRITERIA

An important aspect of stratification dynamics lies in the changes in allocation criteria that result over time from shifting state policies, the topic of our next set of analyses and a major focus of our research. In considering this topic, we stress that as a consequence of revisions in state policies, the prestige and other social rewards of occupations and organizations in China varied across historical periods, as we described earlier. We should bear these contextual variations in mind in interpreting the statistical results.

TABLE 3: Maximum Likelihood Estimates of Piecewise Exponential Models of Occupational Entry in Three Historical Periods: 1949-1965, 1966-1979, and 1980-1993

	Cadre			
Covariates	All Years	1949-65	1966-79	1980-93
Intercept (Model A)[a]		-3.792**	-4.169**	-4.593**
Intercept (Model B)[b]		-3.508**	-4.315**	-8.308+
Education				
Elementary	-1.329**	-1.491**	-1.592	-.250
Senior high	.365	.163	.455	3.847
College	.818**	-.458	1.837**	5.300
Family background				
Cadre	-.120	.192	-.704	.182
"Distrusted"	-.195	.039	-.574	.457
Age	.045*	.092**	.011	.026
Female	-.953**	-1.341**	-.685+	-.756
Beijing	.601**	.802*	.563	-.054
Overall rate per year	.016	.018	.014	.016
Number of events	103	50	36	17

	Professional			
Covariates	All Years	1949-65	1966-79	1980-93
Intercept (Model A)[a]		-5.259**	-5.643**	-5.215**
Intercept (Model B)[b]		-5.308**	-6.119**	-4.793**
Education				
Elementary	-.848	-.525	-4.373	-4.629
Senior high	1.828**	2.140**	1.570**	1.616*
College	3.104**	2.875**	3.308**	3.150**
Family background				
Cadre	.117	.236	-.068	.123
"Distrusted"	.155	.197	.694	-.366
Age	.012	.077**	-.035	.016
Female	.237	.099	.824*	.005
Beijing	.043	.088	.237	-.260
Overall rate per year	.025	.024	.013	.060
Number of events	166	67	35	64

Number of individuals = 980

[a] Model A has time-varying intercepts, which are listed under the historical period for which they apply. Effects of covariates are time-invariant and are listed under "All Years."
[b] Model B has time-varying effects of covariates, as well as time-varying intercepts. Time-varying parameters are listed under the historical period for which they apply.

+ p < .10 * p < .05 ** p < .01

TABLE 3: Maximum Likelihood Estimates of Piecewise Exponential Models of Occupational Entry in Three Historical Periods: 1949-1965, 1966-1979, and 1980-1993 (Continued)

	Worker			
Covariates	All Years	1949-65	1966-79	1980-93
Intercept (Model A)[a]		-2.194**	-2.077**	-1.702**
Intercept (Model B)[b]		-2.260**	-1.844**	-1.848**
Education				
Elementary	-.394**	-.193	-1.212**	-1.567+
Senior High	.240*	-.424+	.183	.703**
College	-2.196**	-3.906**	-1.156*	-2.301**
Family background				
Cadre	-.798**	-.708	-.601*	-1.082**
"Distrusted"	-.722**	-.451*	-.737**	-1.441**
Age	.028**	.046**	.063**	.012
Female	-.066	-.083	-.046	-.157
Beijing	-.148	.075	-.475**	-.037
Overall rate per year	.082	.066	.089	.106
Number of events	533	185	236	112

	Farmer			
Covariates	All Years	1949-65	1966-79	1980-93
Intercept (Model A)[a]				
Intercept (Model B)[b]			-2.791**	
Education				
Elementary			-2.290**	
Senior High			-.490*	
College			-1.422+	
Family background				
Cadre			.533*	
"Distrusted"			.389	
Age			.030	
Female			-.007	
Beijing			-.182	
Overall rate per year	.018	.002	.004	.000
Number of individuals	118	7	110	1

+ $p < .10$ * $p < .05$ ** $p < .01$

As we indicated in describing our methods, to examine changes in allocation criteria across historical periods, we estimated piecewise exponential models of entry into first occupation, first organization, and the Communist Party. For each outcome, we first estimated a model that constrains the covariates to have the same effects in all three historical periods but lets the intercept vary across periods. This model, denoted model A, provides the baseline for examining our argument that the effects of the covariates (i.e., allocation criteria) vary across the three periods. Our argument that the criteria do change over time is represented by a model in which effects of the covariates as well as the intercepts can vary across the historical periods; it is denoted model B. Note that time variation in the values of variables and time variation in their effects are conceptually different; the former is a feature of the data in all of our analyses.

Because we consider multiple destinations for each outcome in each period, the number of transitions to a given destination in a given period is small, except for workers and state-owned firms, which are the most prevalent occupation and type of organization, respectively. Consequently, the statistical power for testing hypotheses about the effect of a single covariate in a single period is low for occupations other than worker and for types of organizations other than state-owned firms. Hence, we consider the magnitude and sign of effects of theoretical interest as well as their statistical significance. We report the result of a one-tailed test when we have a definite hypothesis about the sign of a covariate's effect, but we mainly use two-tailed tests since most of our arguments admit the possibility that effects actually change sign over time.

Entry into a First Occupation

Table 3 gives the maximum likelihood estimates of the effects of covariates on entry into a first occupation. For each occupation, the first column tells the estimates for model A, and the next three columns give the period-specific estimates for model B.

Both model A and model B significantly improve upon a null model with no covariates and intercepts that can vary across the three historical periods; the probability levels for the likelihood ratio χ^2 statistics for these tests are well below .00001. More importantly, the likelihood ratio test comparing model B against model A is 495.0 with 24 degrees of freedom, which also has a *p* level well below .00001.[10] Thus, we can reject the model of time-invariant effects of covariates on the rate of entering a first occupation with considerable confidence. In substantive terms, this means that the criteria used to allocate labor market entrants to their first occupation do vary across the three historical periods, quite aside from the send-down policy that assigned many urban youth to work as farmers during the Cultural Revolution.

Becoming a Cadre. Consider first the pattern of effects in all years, which gives a kind of "average" effect during the 1949-93 period. Relative to a junior high education (the reference category), having a college education has a significant positive effect on the rate of becoming a cadre, whereas having an elementary or no education has a significant negative effect. The rate of entry into this high-status occupation is significantly lower for females and significantly higher for residents of Beijing; it also rises significantly as age increases.[11]

Central to our theoretical concerns, however, are variations across historical periods. Notice first that the intercept steadily and markedly declines over time. The intercept is the log of the rate for the reference category: 20-year-old, male residents of Shanghai with a junior high education and a working class family background. The decline in the intercepts, especially seen in model B, means that over time, cadre are decreasingly chosen from youth with a working-class family background and a junior high education.

The variation in the effects of education over time are especially relevant to our arguments. The general pattern seen in Table 3 is one of positive and increasing effects of having a college education and of negative effects of having an elementary school or no education. The significant, positive effects of having a college education during 1966-79 seem at odds with the discriminatory policy against intellectuals during the Cultural Revolution. We suspect that this result may be due to the favorable policy toward the "Red Guards," many of whom were college graduates. One cannot put much confidence in the small negative effect of an elementary or no education in the last period because only 6% of the observations at risk are in this educational category in that period, whereas 40% are in this category in the first period. The effects of a college or senior high education are large in the 1980-93 period but do not differ significantly from zero. The lack of statistical significance reflects the very small number of moves directly to the cadre occupation in this period, which is partly due to relatively few people in our sample who are at risk of entering the labor force during this period.[12]

Family background indicates individuals' political and cultural capital. The results in Table 3 indicate that those with a "distrusted" or cadre family background have the lowest rate of entry into the cadre occupation during the Cultural Revolution. Although these effects are not statistically significant, which is at least partly due to the few cases in these categories, their effects are consistent with the dominant state policies in force during this historical period. During the Cultural Revolution, older generations of cadres were purged, and a cadre family background became a political liability rather than a political asset. A "distrusted" family background was also especially stigmatized during this period.

The general impression of patterns of entry into the cadre occupation is that by 1980-93, the most important criterion for selection into this occupation is having a higher education. Again, this pattern is consistent with the new cadre policies that emphasize educational qualifications during the reform era.[13]

Becoming a Professional. Education is also very important for entry into a professional occupation. This finding is not so unusual; the same pattern occurs in industrialized market societies. Table 3 shows that having a higher educational level significantly increases the rate of becoming a professional in all three periods. In the first period, having a senior high education and a college education do not seem to be very differentiated. However, over time, having a junior and senior high school education become more similar, and having a college education makes a greater difference. In part this result probably reflects the changing educational composition: Over time, the pool of highly educated youths becomes much larger and the primary source of professionals. Correspondingly, the effect of an elementary school or no education becomes

increasingly negative — the other side of the same coin. The increasing differences in impact of educational levels over time may also mean that the state increasingly relies on educational credentialing in allocation of youth to their first occupation.

With regard to family background, those with a "distrusted" family background have a higher rate of becoming a professional before and during the Cultural Revolution. We interpret this as indicating that such families were better equipped with human and cultural capital to prepare their children for professional jobs. However, given that professionals were stigmatized during the Cultural Revolution, one should not interpret becoming a professional as clearly advantageous. Lastly, those with a cadre family background have a slightly lower rate of becoming a professional during the Cultural Revolution, when they were more likely to become a farmer.

Becoming a Worker. Overall, education affects the rate of becoming a worker in a way that is basically what we had expected. Having an elementary or no education lowers the rate (relative to a junior high level); college graduates also have a lower rate than those with a junior high education, reflecting the fact that college graduates are mainly targeted on cadre and professional occupations. Having a senior high education has a significant negative effect on the rate of becoming a worker in 1949-65 but a significant positive effect in 1980-93. The reversal of sign across these periods should be interpreted in light of changes in the declining effect of having a senior high education on the rate of entry into professional occupations. It signals a reallocation of senior high graduates from professionals in the early period to workers in the 1980-93 period.

Those with a "distrusted" or cadre family background have a significantly lower rate of becoming a worker in almost every period. The effects are especially large and negative during the 1980-93 period, suggesting that by this period, those with either kind of family background are more likely to enter other occupations. It may well indicate that the worker occupation has lost its high prestige by the 1980-93 period and that children from families who are privileged in the new or old systems avoid becoming workers, ceteris paribus. It should be kept in mind that during the Cultural Revolution, workers were elevated to a high status under Mao's famous banner, "The working class is the leading class." A lower rate of entry into a worker occupation during the Cultural Revolution often meant being forced to the countryside.

Becoming a Farmer. Since becoming a farmer resulted mainly from the radical state policies during the Cultural Revolution, there were insufficient cases to estimate a multivariate model of the rate of becoming a farmer except in the 1966-79 period. Here we find the most striking evidence of inconsistent effects of human and political capital on individual life chances. Every educational level has a significantly lower rate of becoming a farmer than the reference category — a junior high education. Those with elementary school or no education and college graduates had the lowest rates of becoming a farmer. Thus, the send-down policy especially affected the life chances of youth with a junior high education. We should keep in mind, however, that the radical state policies in force at that time also lowered the ceiling on the highest level of education that youth could attain.

Those with a cadre family background had a significantly higher rate of becoming a farmer than those having a working class family background. Those with a "distrusted" family background also had a higher rate of becoming a farmer ($p < .10$ using a one-tail test). Substantively, these results indicate that state policies redrew the boundary between "us" and "them" and thereby drastically altered the life chances of those with the previously privileged cadre family background. This shows that the nature of political capital in urban China varied over time.

Entry into a First Organization

We now turn to the analysis of entry into a first organization, considering especially how the criteria used to allocate youth to a first organization vary across the three historical periods. Our findings are reported in Table 4. In general, the patterns parallel those found in the analysis of entry into a first occupation shown in Table 3. This is not surprising because occupation and organization are strongly, though imperfectly, associated. Many cadres work in government agencies; many professionals work in public organizations; workers are usually in state-owned or collective firms; and virtually all farmers are in villages. Indeed farming and being a villager are so closely associated that we do not report results for entry into a village organization in Table 4 — the results for entering a village are almost identical to those for becoming a farmer that are reported in Table 3.

As in the case of our analyses of entry into a first occupation, both models A and B significantly improve upon a null model with no covariates and intercepts that can vary across the three historical periods; the probability levels for these tests are well below .00001. Again, the more interesting test is the comparison of model B and model A. The likelihood ratio χ^2 value for this test is 141.9 with 64 degrees of freedom, which is significant at the .00001 level.[14] Hence, there is strong evidence that the criteria used to allocate youths to their first organizations varied across these three historical periods.

Entry into a Government Agency. Interestingly, the effects of higher education on entry into a government agency are comparatively small and statistically insignificant during the 1949-65 period but larger and statistically significant (for a high school education) during the Cultural Revolution. This anomaly may be because the purge of cadres in government agencies during the Cultural Revolution led to the recruitment into government of those "Red Guards" who graduated from high school or college since political criteria (e.g., loyalty to the Party) were used to recruit college students during this period. By the 1980-93 period, having a college education has a significant positive effect, indicating that educational credentialing is increasingly used to staff positions in government agencies.

The rate of entering a government agency was higher for those with a cadre family background before the Cultural Revolution but significantly lower during the Cultural Revolution. During this period, those with a cadre family background had a significantly higher rate of becoming a farmer, which is equivalent to a village organization, as we said before. So, to a considerable extent, youth with a cadre family background were redirected from government agencies to villages during this period. Those with a "distrusted" family

TABLE 4: Maximum Likelihood Estimates of Piecewise Exponential Models of Organizational Entry in Three Historical Periods, 1949-1965, 1966-1979, and 1980-1993

		Public		
Covariates	All Years	1949-65	1966-79	1980-93
Intercept (Model A)[a]		-4.527**	-5.108**	-4.720**
Intercept (Model B)[b]		-4.303**	-5.593**	-4.782**
Education				
Elementary	-1.040**	-1.071*	-1.558	-7.941
Senior High	1.024**	.953**	.913*	1.297*
College	2.007**	1.373**	2.156**	2.654**
Family background				
Cadre	.122	.411	.373	-.243
"Distrusted"	-.264	-.115	.214	-1.027
Age	.026*	.092**	-.004	.023
Female	.652**	.297	1.305**	.729*
Beijing	.163	.406	-.034	-.086
Overall rate per year	.025	.025	.014	.051
Number of events	163	71	38	54

		Government		
Covariates	All Years	1949-65	1966-79	1980-93
Intercept (Model A)[a]		-3.960**	-4.397**	-4.354**
Intercept (Model B)[b]		-3.842**	-4.233**	-4.904**
Education				
Elementary	-.554+	-.458	-.803	-8.138
Senior high	.311	.092	.779*	-.144
College	.644*	-.314	.900	1.338*
Family background				
Cadre	-.194	.173	-1.293+	-.190
"Distrusted"	-.255	.035	-.794	-.576
Age	.052**	.071*	.072**	.038
Female	-.659**	-1.242**	-.870*	.730
Beijing	.796**	1.069**	.647+	.518
Overall rate per year	.018	.019	.013	.024
Number of events	115	55	35	25

Number of individuals = 980

[a] Model A has time-varying intercepts, which are listed under the historical period for which they apply. Effects of covariates are time-invariant and are listed under "All Years."

[b] Model B has time-varying effects of covariates, as well as time-varying intercepts. Time-varying parameters are listed under the historical period for which they apply.

TABLE 4: Maximum Likelihood Estimates of Piecewise Exponential Models of Organizational Entry in Three Historical Periods, 1949-1965, 1966-1979, and 1980-1993 (Continued)

	State Firm			
Covariates	All Years	1949-65	1966-79	1980-93
Intercept (Model A)[a]		-2.422**	-2.367**	-1.854**
Intercept (Model B)[b]		-2.471**	-2.122**	-2.185**
Education				
Elementary	-.492**	-.348+	-1.501**	-1.535
Senior High	.389**	-.113	.187	1.156**
College	-.088	-.516	.255	.256
Family background				
Cadre	-.738**	-1.456*	-.760*	-.630**
"Distrusted"	-.455**	-.213	-.532*	-.857*
Age	.018*	.045**	.039**	.009
Female	-.195+	-.059	-.166	-.496*
Beijing	-.267**	-.103	-.498**	-.328
Overall rate per year	.068	.054	.068	.107
Number of events	446	153	180	113

	Collective Firm			
Covariates	All Years	1949-65	1966-79	1980-93
Intercept (Model A)[a]		-4.113**	-3.588**	-3.391**
Intercept (Model B)[b]		-4.262**	-3.477**	-3.428**
Education				
Elementary	-.295	.477	-.951+	-.921
Senior High	.099	-.188	.314	-.449
College	-2.326*	-6.678	-1.008	-6.813
Family background				
Cadre	-.789*	.537	-.607	-1.870*
"Distrusted"	-.876*	-.447	-1.166*	-1.195
Age	.038**	.040	.085**	-.019
Female	.096	-.412	.144	.514
Beijing	-.218	-.233	-.522+	.420
Overall rate per year	.015	.010	.020	.020
Number of events	101	28	52	21

Number of individuals = 980

+ $p < .10$ * $p < .05$ ** $p < .01$

background have a lower rate of entering government agencies in both the Cultural Revolution and the reform era. We suspect that many of them entered the small private sector.[15]

Entry into a Public Organization. Since professionals are concentrated in public organizations (e.g., research and educational institutions, hospitals, and mass media), the patterns for entering a public organization are, not surprisingly, very similar to those for entry into a professional occupation. Having a college or senior high education increases the rate of entering a public organization, and these effects tend to become larger over time. Contrarily, having an elementary school education or no education lowers the rate of entering a public organization, and this effect also tends to become larger (i.e., more negative) over time.

Interestingly, those with a "distrusted" family background have a higher rate of entering public organizations during the Cultural Revolution. We think that this pattern was mainly due to cultural capital from their family's prior training in the professions. Again, it is important to remember that during the Cultural Revolution, rewards in public organizations were not greater than those in state-owned firms.

Entry into a State-owned or Collective Firm. We find occasional significant effects of education on entry into state-owned and collective firms prior to the reform era, but in contrast to the other kinds of organizations and all occupations, educational level seems to be used less or at least less consistently — in allocating youth to these kinds of organizations.

During the Cultural Revolution, those with a "distrusted" family background had significantly lower rates of entry into state-owned and collective firms, which provided urban job opportunities. Those with a cadre family background have similar patterns.

Summary. Over time, patterns of entry into a first occupation and organization have changed substantially, but not in a straightforward way. During the recent reform era as compared to earlier periods, educational attainment becomes much more important for entry into high-status organizations and occupations. Although this pattern may be partly due to changing mechanisms of stratification accompanying marketization, it is also consistent with new state policies that emphasize educational qualifications in job allocation.

The effects of family background are different during the reform era: Those with a cadre family background have a significantly lower rate of becoming a worker and of entering state-owned and collective firms, suggesting that they again have the political capital to obtain more favorable jobs. Those with a "distrusted" family background have a lower rate of entering all types of organizations during the reform era, suggesting that they are more likely to enter one of the types of organizations we were unable to study, in particular, the private sector. In any case, the effects of class background appear to persist even in the recent economic era, but in ways different from previous periods.

Although the relatively small number of events has constrained our ability to perform powerful statistical tests of the effects of specific covariates, the directions of effects for the theoretically relevant covariates are quite consistent

with the oral histories of those living through the turbulent history of contemporary China.

Entry into the Communist Party

Next we examine the effects on entry into the Communist Party of an individual's human capital, family background, and own structural position in the occupational and organizational hierarchies. Communist Party membership serves as an indicator of acquired political capital because it is a key political asset for upward mobility in communist societies (see S. Szelényi 1987 on Hungary and Walder 1995 on China). In such societies, membership in the Communist Party is often a prerequisite to promotion into bureaucratic positions, especially ones at higher levels. Examining Party membership also has the advantage that information on state policies about Party recruitment is widely available; hence, results can be interpreted more definitively.

Twenty-three percent of the respondents in our sample were Party members at the time of the 1993 interview. This seemingly large percentage may reflect the concentration of cadre and high-status personnel in Beijing and Shanghai, where we collected the data. Nevertheless, the rate of entry into the Communist Party is low in every period, as we mentioned in our discussion of Table 1. These two facts are not inconsistent because a person tends to join the Party for life.

Our archival research found that state policies on recruitment into the Party shifted over time. In the years shortly after 1949, class background was stressed. After the 1957 Anti-Rightist Campaign, intellectuals were treated with suspicion for recruitment into the Party. In 1964, Party members comprised 15% of those working in the industrial and transportation sectors, 15% in the public and educational sectors, but only 2% of college students (Zhao 1987). In the mid-1960s, a new policy emphasizing intellectuals led to a sharp increase in their percentage in the Communist Party. Policy shifted once more during the Cultural Revolution, again favoring those with a working class background. In the post-Mao era, the leadership again stressed recruitment of intellectuals into the Party. Consequently, we expect to find that effects of human capital, family background, and structural position on recruitment into the Communist Party vary over time.

In our analyses of entry into the Communist Party, we used the same indicators of human and political capital and the same control variables (gender, place, and age) as in our analyses of entry into the labor force. Structural position is represented by time-varying (0-1) indicators of occupation and organizational type. For occupation we distinguish among cadre, professionals, soldiers, and others (workers are the predominant group in this residual category). For organizational type, we distinguish among government agencies, public organizations, collective firms, and others (state-owned firms are the predominant group in this residual category). We also include a time-varying (0-1) indicator of whether the person has not yet entered the labor force to avoid confounding the reference categories of occupation and organization with having no labor force status at all. Table 5 reports the results of our analyses.

First, we note that both model A and model B significantly improve upon a null model in which the covariates have no effects ($p < .00001$). In addition, a

TABLE 5: Maximum Likelihood Estimates of Piecewise Exponential Models of Entry into the Communist Party in Three Historical Periods, 1949-1965, 1966-1979, and 1980-1993

Covariates	All Years	1949-65	1966-79	1980-93
Intercept (Model A)[a]		-4.756**	-4.999**	-5.285**
Intercept (Model B)[b]		-3.374**	-5.610**	-5.677**
Education				
Elementary	-.425+	-.898*	-.261	-.594
Senior High	-.238	-.249	-.222	-.097
College	.094	-.219	-.645+	.781*
Family background				
Cadre	.124	-.034	.281	.340
"Distrusted"	-.019	.060	.011	-.045
Occupation				
Cadre	1.644**	.724	1.697**	1.868**
Professional	.106	-1.374+	.380	.286
Soldier	2.514**	.297	4.293**	1.335
Type of organization				
Government agency	.219	.995*	-.317	.333
Public organization	.153	.140	.551	-.381
Collective firm	-.565	-.874	-.163	-.685
Not yet entered labor force	-.618*	-1.673**	-.534	.015
Age	.003	-.082*	.023	.018
Female	-.516**	-.747*	-.383	-.658*
Beijing	.657**	.493	.877**	.411
Number of events	194	51	73	70
Likelihood ratio χ^2				
Model A vs. baseline	254.6**			
Degrees of freedom	15			
Model B vs. Model A			72.5**	
Degrees of freedom			30	
Number of individuals = 974				

[a] Model A has time-varying intercepts, which are listed under the historical period for which they apply. Effects of covariates are time-invariant and are listed under "All Years."

[b] Model B has time-varying effects of covariates, as well as time-varying intercepts. Time-varying parameters are listed under the historical period for which they apply.

+ $p < .10$ * $p < .05$ ** $p < .01$

test of model B versus model A has a likelihood ratio χ^2 value of 72.5 with 30 degrees of freedom, which is statistically significant at the .0001 level. Thus, there is clear evidence of time variation in effects of the covariates on entry into the Communist Party.

Both structural position and human capital have some time-varying effects on Party recruitment. With regard to organizational type, working in a government agency increases the rate of joining the Party relative to the reference category (predominantly, working in a state-owned firm) in 1949-65, but it has an insignificant effect in subsequent periods. Working in a collective firm has a negative effect in every period, but it never differs significantly from the reference category. In contrast, the rate of joining the Party is significantly higher ($p < .10$ using a one-tail test) for those working in a public organization than in a state-owned firm only during the Cultural Revolution (1966-79). This finding probably reflects the politicization of the public sector (e.g., educational institutions, the media) during that period.

As for occupation, consistent with state policy, professionals had a significantly lower rate of becoming a Party member during the first period as compared with the reference category (predominantly workers). The rate of joining the Communist Party is higher for the cadre category than for the reference category in every period, though the difference in the first period is not statistically significant nor as great as in the later periods. This pattern suggests that cadres were recruited into the Party in every period, though especially after the early period. The consistently positive effect of cadre status here may, however, result from time variation in the criteria used to select cadres, which have varied appreciably over time. We include having a military occupation (being a soldier) in this analysis because joining the army is widely believed to facilitate upward mobility by improving chances of joining the Party (Xiao 1989). Indeed, the effect of being a soldier on the rate of joining the Party is very large and statistically significant during the Cultural Revolution. It is positive but not very large during the early period, and it is large but insignificant in the recent reform era. Very few individuals in our data were soldiers in any period so the statistical power of this test is low. The impact of being a soldier must be very large to achieve statistical significance when the number of cases in the category is small.

Education also shows some time-varying effects. Before the Cultural Revolution, a junior high school education (the reference category) is associated with the highest rate of joining the party; differences based on education are not statistically significant, however. The pattern during the Cultural Revolution is similar, except that those having college education had the lowest rate of joining the Party ($p < .10$). These patterns are consistent with policies in the Party at that time, which emphasized political loyalty more than competence. In the post-Mao era, the situation reverses: College graduates have a significantly ($p < .01$) higher rate of joining the Party than those with a junior high education (and presumably the other educational levels, too, since they have even lower rates of joining the Party than those with a junior high school education). Clearly, human capital has different political implications in China in different historical periods.

Family background does not have a statistically significant effect on joining the Party. This result may be the consequence of prior selection into an occupation and organization based on family background.

Clearly, as far as joining the Communist Party is concerned, the effects of structural locations depend on state policies at a particular time. It would be misleading to discuss the effects of structural position without considering the particular historical context. Given our limited sample size, we are unable to represent all policy-induced variations. However, our findings are consistent with the general patterns of Party recruitment policies reported in official documents.

SUMMARY

Our findings presented above provide empirical evidence for our arguments about stratification dynamics in China.

Varying Opportunities

Educational and work opportunities are structured by the socialist state, and they are therefore highly sensitive to shifts in state policies and have changed substantially over time since the onset of communist rule in China. Access to a college education, for instance, was severely limited during the Cultural Revolution (1966-79). During this period our data show large-scale movements to low-status occupations, to low-status organizations, and to villages. These declining opportunities were a direct product of radical state policies at that time. But we also observe substantial changes between the 1949-65 and 1980- 93 periods in labor force opportunities and in access to membership in the Communist Party. In light of these findings, the timing of entry into the educational system and into the labor force is clearly a critical factor affecting a person's position in the social structure.

Time-Varying Effects of Human and Political Capital

We also found support for our argument that effects of human and political capital (i.e., education and family background) on life chances vary over time. In the case of entry into a first occupation, a first organization, and the Communist Party, the magnitude of these effects, and sometimes their direction, vary in different historical periods. During the Cultural Revolution, a cadre family background significantly lowered the rate of entering a government agency and raised the rate of becoming a farmer (and moving to a village); however, in other periods, a cadre family background tends to have small but positive effects on entering a high-status occupation. Education shows analogous time-varying effects, too.

Time-Varying Effects of Structural Position

We argued that the effects of occupational and organizational position also vary over time. Our findings concerning the effects of these positions on the rate of joining the Communist Party give some pertinent evidence. The effects of types of organization on the rate of joining the Party vary across the three historical periods. The effect of a cadre occupation is consistently positive over time, but

this may be because the criteria used to select cadre vary over time. Still, it seems clear that an association between a cadre occupation and membership in the Communist Party becomes thoroughly institutionalized by the onset of the Cultural Revolution.

Since most respondents in our sample did not shift across occupations or organizations after their first entry into the labor force (except for sent-down youth, most of whom returned to urban areas after 1979), we believe that the shifting policy-induced patterns of entry into the labor force, such as those we found, have a lasting effects over individuals' life course.[17]

Discussion: The State and Life Chances

Obviously our findings are highly bounded contextually and they should be, given our emphasis on the importance of historical context and political dynamics associated with twists and turns in state policies. However, our arguments have broader implications for the comparative study of stratification.

Our central argument pertains to the time-varying effects of human, social, and political capital and of structural position on individual life chances. In the Chinese context, such time variation was particularly salient during the Cultural Revolution. An obvious question is: Does this period represent an anomaly — a large random shock to the social system that is highly unlikely to occur again? If this is the case, our findings have limited generalizability.

Although the Cultural Revolution represents an extreme case in China's political history, Chinese macropolitical processes are much more dynamic, dramatic, and unpredictable than portrayed by the simplified historical periods used in our empirical analyses. For instance, the Anti-Rightist Campaign of 1957 purged over 550,000 intellectuals who were rehabilitated only twenty years later. In 1960, the Communist Party sent down over one million cadres to the lower levels — about 15% of the total number of cadres then (Zhao 1987). During the Cultural Revolution, 17% of cadres and 75% of high-level officials were interrogated and purged. Including their family members who also bore the consequences of their downward mobility, it is estimated that a hundred million people suffered during this period (Chen & Liu 1991). Indeed, fluctuating state policies are a regularity of political life in China. Because we focused on such life events as first occupation, first organization, and joining the Communist Party, our analyses have not treated the more prevalent and more subtle forms of policy-induced variations in life chances that happen within a given occupation and organization.

One may be skeptical about the importance that we attribute to political dynamics and think that the patterns we have reported are unique to Communist China. Although we have examined China since 1949, our discussion is motivated by broad differences between market and redistributive economies and by the political dynamics intrinsic to state socialism. Political dynamics affect individual life chances in the ways that we have discussed only when politics are closely linked with everyday life. We think that our arguments have implications for all societies where political authority attempts to manipulate individual life chances directly. The processes of resource transfers may take

different forms, depending on the specific institutional structure of political authority, but the effects of political dynamics on individual life chances are — we believe — inevitable in such societies.

Attention to stratification dynamics raises both theoretical and methodological issues in the comparative study of stratification. Theories not only need to give attention to structural conditions but also take seriously evolving political processes, especially shifts in state policies and other noteworthy political events. Our analyses suggest that, to understand stratification in urban China in particular and in state socialist societies in general, one should pay attention not only to the unique institutional structure of redistributive economies, but more importantly to macropolitical processes and state policy shifts. Stressing individual-level resources (social origins, human capital) and lateral group competition, as is common for industrialized market societies, is problematic: Individuals do make choices, and resources do facilitate upward mobility. But the extent to which choices are available and the value of various resources cannot be understood without an adequate knowledge of the macropolitical environment.

We distinguish sharply between stratification processes that follow a political logic and those governed largely by a market logic. The emergence of a central state and political mechanisms of stratification, which is characteristic of (but not exclusive to) state socialist societies, inevitably links state policies with drastic social dislocations. The rise of the welfare state has also enhanced the active role of political authority in stratification processes in industrialized market societies. Governmental social policies in the U.S. have, for example, played a significant role in equalizing opportunities. According to DiPrete and Grusky (1990): "The most important 'independent forces' in advanced industrial systems may be the political policies and programs undertaken by the state itself" (109). As the modern state expands and penetrates civil life, we expect channels of resource transfer to be altered and the effects of state policies on individual life chances to increase. Given the growing role of the state in the labor markets of industrialized market societies, the Chinese experience has implications far beyond China and other state socialist societies.

Notes

1. Whether one should stress the "family resemblance" of all such regimes or the nation-specific departures from a common regime is the subject of much debate. but viewed broadly, the family resemblance is more apparent than national uniqueness.

2. A growing literature emphasizes organizational context and the state in stratification in industrialized market societies, too. See, for example, Althauser and Kalleberg (1981) and Baron and Bielby (1980) on the impact of organizational context and Mayer and Schoepflin (1989) and DiPrete and Grusky (1990) on the role of the state.

3. Potential sample selection bias is another issue. We drew our sample from households that fell in our sampling frame in the two cities. Since our study focuses on urban China, the exclusion of rural areas does not impede our analyses. Random flows of urban residents across urban areas should not cause sample selection bias. One potential problem is the exclusion of those who left the urban areas before our data collection began. For instance, some "sent down" youth remain on farms even today. To the best of our knowledge, however, the

percentage who did not return to urban areas after the send-down policy was reversed is very low (Deng 1993:351) and therefore should not appreciably affect our analyses.

4. A few other types of organizations, such as private firms, exist in some historical periods but occur too infrequently in our data to be distinguished in our analyses.

5. The main exception occurs when a parent holds a "cadre" position after 1949, which takes precedence over class position of parents and grandparents before 1949.

6. In these data, the number of events in any given historical period is too small to have a reasonable level of statistical power if more and narrower historical periods are used. However, at the suggestion of one reviewer, we explored five somewhat narrower periods in our analysis of entry into the Communist Party. The general pattern of results was similar to that obtained with the three broad historical periods. For consistency, we report below only results for the three broad historical periods.

7. A better picture of educational opportunities would be given by the number enrolled in a given educational level divided by the number of graduates from the level below. But we found data only on the annual number of graduates at each educational level. Therefore, Figure 1 only indirectly indicates time variation in educational opportunities.

8. Being a farmer is almost perfectly correlated with a village type of organization.

9. The term *population at risk* or *risk set* refers to the set of individuals who can experience some event. In general, the members of the risk set vary over time. For example, at any given time, only non-Party members who are at least 18 years old are eligible to join the Communist Party, and only people who are old enough to work but who have not yet started to work are at risk of entering a first occupation and first organization.

10. This test treats farmers as censored because so few people enter this occupation except during the Cultural Revolution. Hence, including farmers would give a significant difference over time, even if nothing else changed. The test reported here is more conservative because it pertains to changes over time in effects of covariates on entering only the cadre, professional, and worker occupations.

11. We do not comment on effects of the control variables (gender, place, and age) except when they seem to be of particular interest.

12. The number of people at risk of entering the labor force is 195 in 1980-93 but over 500 in the other two periods (see Table 2). The relatively small number in 1980-93 reflects the fact that individuals younger than 25 were not interviewed. This interviewing rule causes the average age and educational levels of those at risk of entering the labor force to be somewhat higher in 1980-93 than would be the case if a lower minimum age had been used to select the sample.

13. We did not include Communist Party membership in the model because we are examining entry into the first occupation and first organization, and labor force entry usually occurs before a person joins the Party.

14. We treat a move to a village organization as censored in this test (as we treated a move to a farmer occupation as censored in the corresponding test in Table 3) to avoid finding a significant effect solely because moves to a village organization occurred almost only during the Cultural Revolution.

15. We treated entry into the private sector as censored in our analyses because the number of events was too small to perform any multivariate analysis of moving to this destination.

16. We note that relatively similar percentages of Party members were found in Tianjin in the 1980s (Lin & Bian 1991; Walder 1995).

17. We plan to examine these effects in future studies. For a preliminary report, see Zhou, Moen & Tuma 1994.

References

Althauser, Robert P., and Arne L. Kalleberg. 1981. "Firms, Occupations and the Structure of Labor Market: A Conceptual Analysis." Pp. 119-49 in *Sociological Perspectives on Labor Markets*, edited by Ivar Berg. Academic Press.

Baron, James N., and William T. Bielby. 1980. "Bringing the Firm Back In: Stratification, Segmentation, and the Organization of Work." *American Sociological Review* 49:454-73.

Bauman, Zygmunt. 1974. "Officialdom and Class: Bases of Inequality in Socialist Society." Pp. 129-48 in *The Social Analysis of Class Structure*, edited by Frank Parkin. Tavistock Publications.

Bian, Yianjie. 1994. *Work and Inequality in Urban China*. SUNY Press.

Blau, Peter M., and Otis Dudley Duncan. 1967. *The American Occupational Structure*. Wiley.

Blau, Peter M., and Danching Ruan. 1990. "Inequality of Opportunity in Urban China and America." *Research on Social Stratification and Mobility* 9:3-32.

Breiger, Ronald L. 1981. "The Social Class Structure of Occupational Mobility." *American Journal of Sociology* 87:578-611.

Burawoy, Michael, and Janos Lukács. 1985. "Mythologies of Work: a Comparison of Firms in State Socialism and Advanced Capitalism." *American Sociological Review* 50:723-37.

Chen, Zhili, and Jie Liu. 1991. *Gongguo Gongchandong Jiansheshi [The History of the Development of the Chinese Communist Party]*. Shanghai: Shanghai People's Publishing House.

China State Statistical Bureau. 1990. *Gesheng, Zizhiqu, Zhixiashi Lishi Tongji Zhiliao Huibian [Historical Statistics of the Provinces, Autonomous Regions, and Metropolises]*. Beijing: China Statistics Press.

Connor, Walter. 1979. *Socialism, Politics, and Equality: Hierarchy and Change in Eastern Europe and the USSR*. Columbia University Press.

Davis, Deborah. 1992a. "Skidding: Downward Mobility Among Children of the Maoist Middle Class." *Modern China* 18:410-37.

____. 1992b. "Job Mobility in Post-Mao Cities: Increases on the Margins." *China Quarterly* 84:1062-85.

Deng, Xian. 1993. *Zhongguo Zhiqingmeng [The Dreams of the Educated Youth in China]*. Beijing: The People's Publishing House.

DiPrete, Thomas, and David B. Grusky. 1990. "Structure and Trend in the Process of Stratification for American Men and Women." *American Journal of Sociology* 96:107-43.

Djilas, Milovan. 1966. *The New Class*. Unwin Books.

Elder, Glen H., Jr. 1974. *Children of the Great Depression*. University of Chicago Press.

Erikson, Robert, and John D. Goldthorpe. 1992. *The Constant Flux*. Clarendon Press.

Featherman, David, and Robert M. Hauser. 1978. *Opportunity and Change*. Academic Press.

Goldthorpe, John H. 1987. *Social Mobility and Class Structure in Modern Britain*. 2d ed. Clarendon Press.

Grusky, David B., and Robert M. Hauser. 1984. "Comparative Social Mobility Revisited: Models of Convergence and Divergence in Sixteen Countries." *American Sociological Review* 49:19-38.

Hauser, Robert M., and David L. Featherman. 1977. *The Process of Stratification*. Academic Press.

Hauser, Robert M., and David B. Grusky. 1988. "Cross-National Variation in Occupational Distributions, Relative Mobility Chances, and Intergenerational Shifts in Occupational Distributions." *American Sociological Review* 53:723-41.

Hout, Michael. 1988. "More Universalism, Less Structural Mobility: The American Occupational Structure in the 1980s." *American Journal of Sociology* 93:1358-400.

Konrád, George, and Ivan Szelényi. 1979. *The Intellectuals on the Road to Class Power*. Harcourt Brace Jovanovich.

Kraus, Richard Curt. 1981. *Class Conflict in Chinese Socialism.* Columbia University Press.

Lee, Hong Yung. 1991. *From Revolutionary Cadres to Party Technocrats in Socialist China.* University of California Press.

Lin, Nan, and Wen Xie. 1988. "Occupational Prestige in Urban China." *American Journal of Sociology* 93:793-832.

Lin, Nan, and Yanjie Bian. 1991. "Getting Ahead in Urban China." *American Journal of Sociology* 97:657-88.

Lindblom, Charles E. 1977. *Politics and Markets: The World's Political-Economic Systems.* Basic Books.

Mayer, Karl Ulrich, and Urs Schoepflin. 1989. "The State and the Life Course." *Annual Review of Sociology* 15:187-209.

Nathan, Andrew J. 1976. "Policy Oscillations in the People's Republic of China." *China Quarterly* 68:720-33.

Nee, Victor. 1989. "A Theory of Market Transition: From Redistribution to Markets in State Socialism." *American Sociological Review* 54:663-81.

____. 1991. "Social Inequalities in Reforming State Socialism: Between Redistribution and Markets in China." *American Sociological Review* 56:267-82.

Parish, William L. 1984. "Destratification in China." Pp. 84-120 in *Class and Social Stratification in Post-Revolution China,* edited by James L. Watson. Cambridge University Press.

Rosen, Stanley. 1982. *Red Guard Factionalism and the Cultural Revolution in Guangzhou (Canton).* Westview Press.

Sorensen, Jesper. 1992. "Locating Class Cleavages in Intergenerational Mobility: Cross-National Commonalities and Variations in Mobility Patterns." *European Sociological Review* 8:267-81.

Stark, David. 1986. "Rethinking Internal Labor Markets: New Insights From a Comparative Perspective." *American Sociological Review* 51:492-504.

Szelényi, Ivan. 1978. "Social Inequalities in State Socialist Redistributive Economies." *International Journal of Comparative Sociology* 19:63-87.

____. 1988. *Socialist Entrepreneurs: Embourgeoisement in Rural Hungary.* University of Wisconsin Press.

Szelényi, Ivan, and Robert Manchin. 1987. "Interrupted Embourgeoisement: Social Background and Life History of Family Agricultural Entrepreneurs in Socialist Hungary." *Research in Social Stratification and Mobility* 8:253-78.

Szelényi, Szonja. 1987. "Social Inequality and Party Membership: Patterns of Recruitment into the Hungarian Socialist Workers' Party." *American Sociological Review* 52:559-73.

Titma, Mikk, and Nancy Brandon Tuma. 1993. "Stratification Research in a Changing World." Pp. 225-54 in *Eastern European Societies at the Threshold of Change,* edited by J. Szmatka, Z. Mach, and J. Mucha. Edwin Mellen Press.

Treiman, Donald J. 1970. "Industrialization and Social Stratification." Pp. 207-34 in *Social Stratification: Research and Theory for the 1970s,* edited by Edward O. Laumann. Bobbs-Merrill.

Tuma, Nancy Brandon, and Michael T. Hannan. 1984. *Social Dynamics: Models and Methods.* Academic Press.

Unger, Jonathan. 1982. *Education under Mao.* Columbia University Press.

Walder, Andrew G. 1986. *Communist Neo-Traditionalism: Work and Authority in Chinese Industry.* University of California Press,

____. 1992. "Property Rights and Stratification in Socialist Redistributive Economies." *American Sociological Review* 57:524-39.

____. 1995. "Career Mobility and the Communist Political Order." *American Sociological Review* 60:309-28.

Weber, Max. 1978. *Economy and Society.* Edited by G. Roth and C. Wittich. University of California Press.

Whyte, Martin King. 1973. "Bureaucracy and Modernization in China: The Maoist Critique." *American Sociological Review* 38:149-63.

____. 1985. "The Politics of Life Chances in the People's Republic of China." Pp. 244-65 in *Power and Policy in the PRC,* edited by Y.M. Shaw. Westview Press.

Whyte, Martin King, and William L. Parish. 1984. *Urban Life in Contemporary China.* University of Chicago Press.

Xiao, Qinfu. 1989. Wuci *Longchao [Five Waves].* Beijing: The People's University Press.

Yamaguchi, Kazuo. 1983. "The Structure of Intergenerational. Occupational Mobility: Generality and Specificity in Resources, Channels, and Barriers." *American Journal of Sociology* 88:718-45.

Zhao, Shenghui. 1987. *Zhongguo Gongchandong Zhuzhishi Gongyao [The History of the Communist Party Organization].* Anhui: Anhui People's Publishing House.

Zhou, Xueguang. 1992. "Lun yiguan jiusi yifang jiuluan" [On Political Cycles in Contemporary China]. Pp. 149-69 in *Dangdai zhongguo de guoj'za yu shehui guanxi [State and Society in Contemporary China],* edited by Xueguang Zhou. Taipei: Guiguan Press.

____. 1993. "Unorganized Interests and Collective Action in Communist China." *American Sociological Review* 58:54-73.

____. 1995. "Partial Reform and the Chinese Bureaucracy in the Post-Mao Era." *Comparative Political Studies* 28:440-68.

Zhou, Xueguang, Phyllis Moen, and Nancy Brandon Tuma. 1994. "The Political Economy of Job Transition in Urban China: 1979-1993." Paper presented at the annual meeting of the American Sociological Association. Los Angeles, Ca.

Decomposing the Intellectuals' Class Power: Conversion of Cultural Capital to Income, Hungary, 1986*

JÓZSEF BÖRÖCZ, *Rutgers University*
CALEB SOUTHWORTH, *University of California, Los Angeles*

Abstract

Konrád and Szelényi posed a powerful critique of state socialism with their formulation of the new class theory suggesting that intellectuals were "on the road to class power." Using a Weberian reading of Bourdieu, we specify empirically one aspect of that critique and examine the statistical effects of cultural capital on income inequalities in Hungary, 1986. While confirming the independent effects of both "education" and "habitus" — potential and realized components of cultural capital — our analysis suggests that certification obtained through formal education is the strongest cultural capital predictor of incomes.

During the years between the end of World War II and 1989, the socialist states of Soviet-controlled Europe claimed to have reorganized society to exclude private ownership of the major means of production. While the political-economic verity of that claim may have been questionable, there was a certain sociological validity to it. Following the Soviet pattern, all socialist states abolished the propertied bourgeoisie both as locus of accumulation and as a socially recognized group.

** Data for this study come from the archives at TÁRKI, the Budapest-based Association for Social Research and Informatics, and were obtained free of charge due to the senior author's affiliation with one of TÁRKI's member institutions. The preparation of this study was supported by grants received from the Center for German and European Studies at the University of California, Berkeley, from OTKA (the National Science Research Fund of Hungary under number T6739), and by further support from the Institute for Political Science of the Hungarian Academy of Science and the University of California at Irvine. Computing facilities were partly provided by the University of California at Irvine. The authors owe special thanks to Ildikó Merkl for providing invaluable help in obtaining the data and Gábor Körösi for some useful initial statistical discussion. We are grateful for detailed comments received from Judit Bodnár, Éva Fodor, David A. Smith, Judith Stepan-Norris, and Max Stevens, and the two anonymous reviewers at Social Forces on earlier drafts. The usual disclaimers concerning the text's remaining weaknesses apply.* *Direct all correspondence to József Böröcz, Department of Sociology, Rutgers University, New Brunswick, NJ 08903-5072 or e-mail:jborocz@rci.rutgers.edu.*

The idea that the abolition of the propertied bourgeoisie did not result in the elimination of the class nature of such societies has been the centerpiece of a powerful critique of state socialist practices (Djilas [1957] 1962b; Gouldner 1979; Konrád & Szelényi [1974] 1979). In its most mature form, the "new class" thesis can be reformulated for our purposes as arguing that cultural capital — possession of which is a marker of intellectuals — replaced economic property as the fundamental structure of social inequality under state socialism.

This article attempts an empirical specification of the new class thesis with respect to income inequalities using survey data from three years before the collapse of state socialism in Hungary. The models presented here sort out the effects of cultural capital and traditional stratification variables on income variation; then the effects of cultural capital are decomposed. The article demonstrates the application of cultural capital to a socialist society and the particular use of the components of cultural capital in a social stratification model of income.

Intellectuals, Cultural Capital, and State Socialism

In the official rhetoric of state socialism, inherited social inequalities were reduced through the redistributive system which allocated resources for private as well as public consumption. The elimination of the social position of the exploiter class, and the ceaseless redistributive action of the socialist state, were supposed to have equalized the relations of distribution. Work on comparative social inequality indicates that during the decades of state socialism levels of social inequality were indeed reduced in comparison to prewar conditions. However, the existence of social disparities and, specifically, income inequalities, cannot be denied.[1] This was especially so toward the late period of state socialism, and particularly in Hungary, one of the state socialist societies that drifted farthest away from the Soviet pattern.

The possibility of intellectuals taking power in a socialist society is not a new idea. Even before the Russian Revolution, Polish-born Bolshevik Jan Machajski argued that socialism was the intelligentsia's aspiration to abolish capitalism and nationalize the means of production in its own interest (Stites 1989). After that revolution, Leon Trotsky ([1933] 1972) attacked what he saw as a "deformed workers' state" and rule by a corrupt bureaucracy instead of the working class. With almost four decades of Soviet and ten years of Yugoslav experience behind him, Milovan Djilas refined Trotsky's criticisms into a theory of "the new class" in which he argued that the ruling bureaucracy had taken on a class character (Djilas [1957] 1962b, 1962a) on the basis of their organizational power: "The bureaucrats in a non-Communist state are officials in modern capitalist economy, while the Communists are something different and new: a new class" (Djilas [1957] 1962b:44).

Finally, in the post-Sputnik Brezhnev years — when groups of humanistic and technocratic intellectuals began to enter the party and government with hopes of rationalizing the state — Konrád and Szelényi developed the idea of the "new class" into a theory of intellectual class interest articulated through a state-run system of rewards for education and technical skill (Konrád & Szelényi

[1974] 1979; Szelényi 1982, 1986-87; see also Gouldner 1979). They argued that increased reliance on experts, combined with the meritocratic educational and political system, provided an environment in which intellectuals gradually came to control the redistributive apparatus of the state, a power that gave them the possibility of disposal in their own interest (Szelényi 1986-87).

An analytical distinction between the propertied bourgeoisie and the cultural bourgeoisie helps explain the evolution of the intellectuals' dominance. In a recent paper, Konrád and Szelényi address a historical characteristic of the belated, semiperipheral development of capitalism in East Central Europe[2] and distinguish between two German terms, *Besitzbürgertum* and *Bildungsbürgertum*, as follows: "we will use the distinction between *Bildungsbürgertum* (the educated middle class, which is bourgeois, or *bürgerlich*, primarily in terms of its habitus, lifestyles, consciousness, or commitment to pluralistic democratic institutions and civil liberties) and *Besitzbürgertum* (the propertied) as a characteristically central European distinction that presumes the openendedness or undetermined nature of the dialectical relationship between these two phenomena" (359). In central Europe, these two aspects of west European bourgeois development were divorced from each other in terms of both space and time, and were located in social groups and behavioral patterns that were distinct in terms of their sociocultural roots, claims of legitimacy, patterns of social networks, and relationships vis-à-vis the state and the world economy. Hence, a key consequence of the transition to state socialism was that the removal of the *Besitzbürgertum* elevated the *Bildungsbürger* practice and its practitioners (some of whom are now called intellectuals) to new heights of social power, prestige, and privilege.

Our analysis is *not* a direct test of Konrád and Szelényi's proposition pertaining to the class nature of state socialism or of the question of whether or not the intellectuals constitute a class. Instead we ask: *If* state socialism is a class society, and *if* the intellectuals occupy a systematically advantageous position in it, then what specific components of cultural capital have positive statistical impact on one specific form of inequality, income levels?

This article first builds a basic stratification model of income. Then it shows the effects of cultural capital on earnings differences under state socialism. Our model uses Bourdieu's (1984) definition of *cultural capital* as "the disposal of taste" or consumption of specific cultural forms that mark people as members of specific classes (1984).

In the north American version of the Weberian tradition, "'status' (*ständische Lage*) shall mean an effective claim to social esteem in terms of positive or negative privileges" (Weber [1920] 1978:305).[3] The components of status are (1) style of life,[4] (2) formal education, and (3) hereditary or occupational prestige (305-6). Cultural capital is Bourdieu's reinterpretation of Weber's theory of status groups. Bourdieu defines *cultural capital* as competence in a society's high-status culture. He argues that it can be broken down into a realized, or *habitus*, and a potential, or *education*, component.[5] The former includes perceptions, appreciation, and specific actions (Bourdieu 1977). Thus, habitus — Bourdieu's conceptually more restrictive equivalent of the "style of life" component of status in Weber — is defined as "a system of lasting, transposable dispositions which, integrating past experiences, functions at every moment as

a matrix of perceptions, appreciations and actions and makes possible the achievement of infinitely diversified tasks, thanks to analogical transfers of schemes permitting the solution of similarly shaped problems" (82-83).

Bourdieu describes education as potential cultural capital because a large part of sociocultural practices are organized by the "dormant" aspects or predispositions contained in individuals' cultural endowments. Denoting not only formal institutional, but also informal aspects of such dormant predispositions, Bourdieu's *education* is a conceptually larger version of Weber's *formal education* (Bourdieu 1986:243-46).

Our analysis follows Bourdieu in operationalizing cultural capital as competence in high-status culture where "high-status culture" is not a measure of specific content, but rather the specific value society places on a given artist, form of expression, knowledge or communication, commodity, or pattern of behavior. We break it down along the two axial dimensions — education and habitus.

Lamont and Lareau (1988) define cultural capital as "institutionalized, i.e., widely shared, high status cultural signals . . . used for social and cultural exclusion" (156). Read from the angle of state socialist class analysis, it becomes clear that Lamont and Lareau need to cast cultural capital as a set of "signals" because they address capitalist society where both cultural capital and economic capital display full and relatively distinct hierarchies of inequality that are not simply reducible to each other. At a minimum, a process of mediation, symbolization, or, as in Lamont and Lareau, "signalling," is necessary between social structures emerging from economic capital and structures of cultural capital.

With the establishment of state socialism, however, private individuals were removed from positions in which they could benefit from exploitation based on economic capital.[6] This empties the property-based dimension of class conflict and opens up new areas as potential dimensions of class formation. While the previous ruling classes could be economically expropriated, their cultural and social practices and tastes could largely endure. With the state as the owner of large-scale capital and the controller of the labor process, formal mechanisms of redistribution lead to the large-scale transformation of the class structure. The hierarchy of the state bureaucracy becomes the basic structure of state socialist society. The internal logic of state organization provides new opportunities for a select group of individuals to use their structural advantages in terms of cultural capital through institutional mechanisms.[7] Instead of being abolished, social exclusion is reorganized and systematized along different lines. Our study examines one particular aspect of this reorganization of exclusion: the conversion of cultural capital into personal earnings.

Our image of the causal mechanism connecting cultural capital and income levels can be summarized as follows: Cultural capital comes in two major forms, education and habitus. Habitus is associated with practices of cultural production and consumption that enable its practitioners to maintain a more detailed, informed, and complex vision of their social world. It is also a source of social respect and prestige. Individuals who are in command of those are better prepared to foresee changes in their environment and have more precise models of the social world available for them. The patterns of taste, judgment, and

distinction that emerge in this process function as cultural reinforcement mechanisms solidifying systematic inclusion and exclusion.

Education is a central marker in this hierarchy. Along with all the advantages of prestige associated with habitus, education confers a further level of advantage in the sense that state enterprises require certain job-specific levels of skills and qualifications to regulate entry and promotion. Furthermore, potential aspects of cultural capital are more fluid, less objectified, and less restricted by existing social patterns of behavior than habitus. Consequently, they are more amenable to complex strategies of advancement in the hierarchical division of labor of this state socialist society. Both habitus and education are expected to contribute to improving the individual's position in the system of distributional advantages, but the effects of education are expected to be stronger and more direct.

Data, Variables and Method

The data come from TÁRKI-III, a general social survey of Hungary executed as a cooperative effort by the Sociology Institute of the Eötvös Loránd University, the Mass Communication Research Center and the Association for Social Research and Informatics in 1986. It uses a three-step stratified sample ($n = 2{,}963$) of the country's adult civilian population.[8]

Table 1 includes the variables, brief descriptions, value labels, means, and standard deviations. The dependent variable of the analysis is the logarithm[9] of the total of all the respondent's self-reported individual incomes expressed in monthly figures in Hungarian Forints.[10] It includes money incomes from work (wages and salaries), social incomes in cash (maternity, retirement or disability benefits, workers' compensation, sick pay, scholarships and fellowships, and aid in cash), and estimates of incomes in cash and in kind from the second economy; for example, second jobs, overtime pay, tips, moonlighting incomes, and revenues from "around the house" farming. However, the income variable excludes social incomes in kind, such as health care, child care, education, housing, and transportation subsidies. While there is an important literature that assesses other sources of inequality under socialism, particularly housing, we see those sources as a separate empirical problem.[11] Our effort is to show that cultural capital matters to income inequality, not that it is the only dimension of social stratification in Hungary. The survey's mean monthly income figure — HUF 6,357.82 per month — is the equivalent of U.S.$138.73 at the annual average official exchange rate for 1986 (KSH 1991), and is almost exactly the same as the estimated national figure of HUF 6,310.80 (U.S.$137.70) provided by the Central Statistical Office for 1986 (KSH 1991). We exclude those cases with no income from our analysis because we have no reason to believe that our model of income determinants would predict zero income. We now detail the baseline measures used to control for income variation and their expected effects and next turn to the cultural capital variables.

Party membership should be a significant contributor to income inequality as advancement to the higher ranks of the occupational hierarchy required party membership. It is coded in two dummy variables: *high party* for those with a

TABLE 1: Description, Value Labels, Means, and Standard Deviations of Dependent and Independent Variables

Variable	Description	Values	Mean	S.D.
Income	Total individual income from all sources (in Forints)		6,357.82	5,057.92
Log income	Logarithm of total income		3.723	.254
High party	Party member with workplace function or higher	0=no, 1=yes	.051	.220
Low party	Party member with less than workplace function	0=no, 1=yes	.091	.287
Age	Age		45.147	16.159
Both sectors	Respondent works in state and nonstate sectors	0=no, 1=yes	.141	.348
Nonstate	Respondent works in nonstate sector only	0=no, 1=yes	.027	.161
Gender	Gender	0=male, 1=female	.525	.499
Work	Respondent actively works for income	0=no, 1=yes	.696	.460
Budapest	Respondent lives in Budapest	0=no, 1=yes	.210	.408
Other city	Respondent lives in other large city	0=no, 1=yes	.348	.486
Basic education	Schooling: complete elementary	0=no, 1=yes	.649	.477
High school	Schooling: completed high-school	0=no, 1=yes	.239	.462
College	Schooling: completed college	0=no, 1=yes	.059	.236
University	Schooling: university diploma	0=no, 1=yes	.049	.216
Number of books	How many books do you have?		310.7	646.7
Read books	How often do you read books?	1=never, 2=sometimes, 3=often	2.174	0.733
Theater	Visits/year to the theater during last year		1.574	3.648
Classical music	Visits/year to classical concert during last year		.390	2.238
Foreign language	Can you speak a foreign language?	1=no, 2=so-so, 3=fluent	1.373	0.686
Trips abroad	How many times have you been abroad?	0-99 or more	4.756	9.520

cell-level function in the workplace, a county-level function, or who served on the central committee; *low party* for simple members. Nonmembership is the implicit point of comparison.

Variables *age* and *gender* are markers of basic socially ascribed traits of the individual. Income should go up as the person becomes older and more capable and down after retirement age. The nonlinear effect of retirement is filtered out by the use of the control variable *work* which marks active labor force participation.[12] Further, gender stratification and systematic gender-based discrimination against women exists in all societies, including state socialist ones.

We use two dichotomous variables to show the respondent's presence in either the *second economy only* or both the first and second economy (*both economies*). The reference category is the majority of the population who participate in the state sector only. Of those who report incomes in our sample, 82.1% are employed only in the state sector, 15.3% are bisectoral, and a mere 2.6% earn incomes solely in the private sector. Participating in the first and second economy are those individuals who draw incomes from the state sector, and who *also* are self-employed, earn tips, conduct farming around the house, receive "gratitude monies" (typically given to doctors), or subcontract from the state. Participating in the second economy exclusively are the small number of respondents who hold no state employment or entitlement; these people are concentrated in agriculture and artisanal self-employment (Böröcz 1993; Róna-Tas 1990).

Finally, two more dummy variables control for the urban/rural contrast and urban primacy: *Budapest* indicates those who live in the capital city and *other city* shows those who live in other settlements officially classified under Hungarian administrative rules as urban.[13] These screen out the effects of cultural opportunities and income earning strategies that exist only in larger metropolitan areas.

Cultural capital is measured via variables organized along the two axes defined by Bourdieu. Formal educational credentials, proficiency in foreign languages, and the number of books the respondent owns measure the education dimension. A series of dummy variables captures different levels of schooling completed, reproducing the emic structure of schooling levels in Hungary, and ranges from complete elementary school education (*basic education*), complete high school (*high school,*) and a college degree (*college*) to a university diploma (*university*). Less than a completed elementary school education is the reference category. The approximate self-reported number of books owned is *number of books*. The ability to speak a foreign language is coded as a three-level, rank-order variable (*foreign language*).

Measures classified under habitus include the number of hours per week spent reading (*read books*); the number of times per year the respondent goes to the theater (*theater*); the number of times per year the respondent attends a classical music concert (*classical music*); and the total number of trips taken abroad during the respondent's life (*trips abroad*).

Measures of certification are expected to be the most fundamental predictor variables in this model. We think certification affects income independently in that most occupations explicitly require specific levels of certification. Parental income/wealth exert some influence on certification, because money can buy

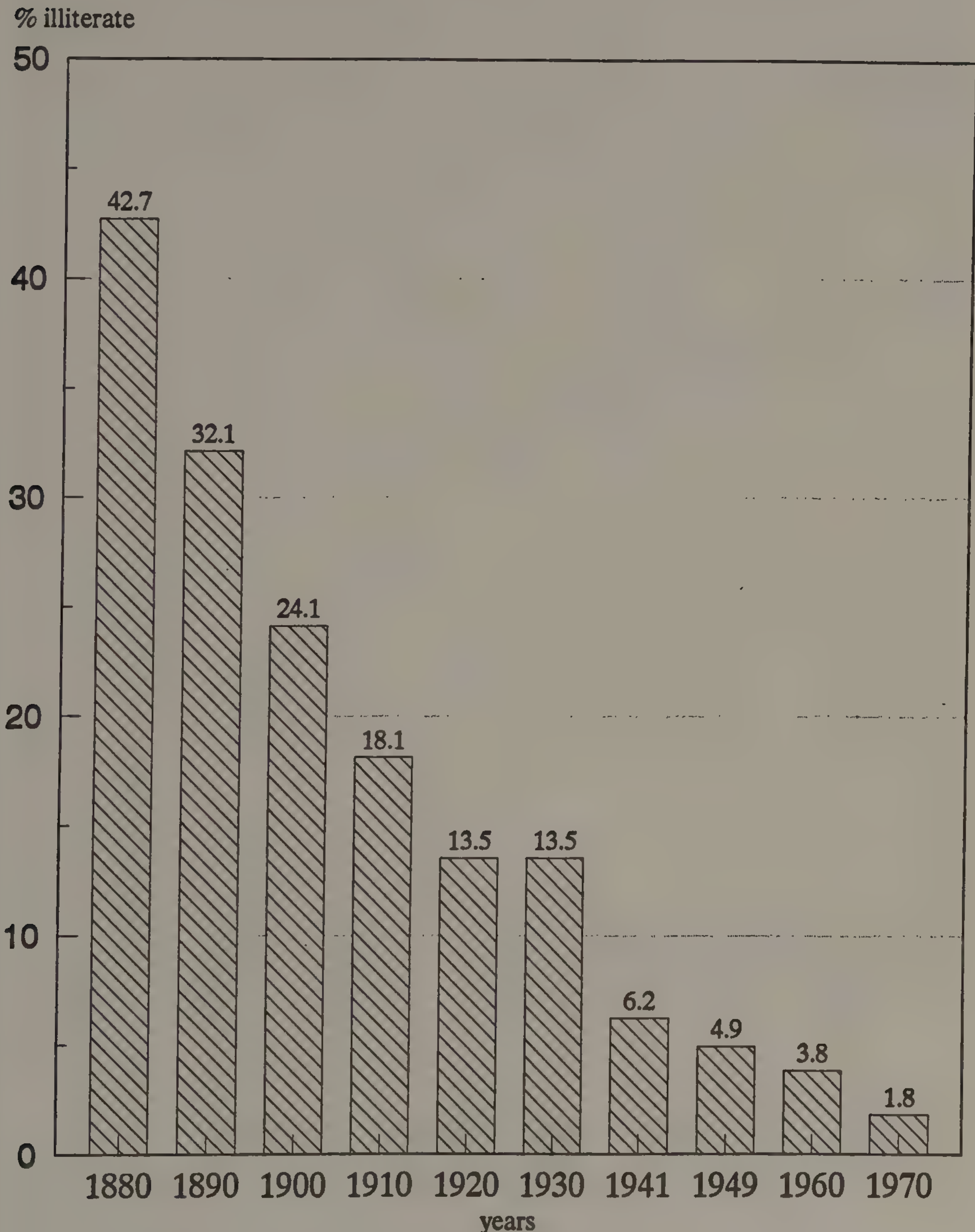
resources that make education easier, but this influence, of course, is much smaller in Hungary than the U.S. where ability to pay is a condition for many educational advancement decisions. In Hungary, educational attainment reflected in certification involves both the drive for economic ends and a particular underlying value system. Individuals' passage through hierarchically structured institutional forms of education exposes them to ideas and endows them with symbolic markers of social prestige and power. Drawing on history, art, and the humanities allows them to envision more possibilities than their immediate, current reality. Success in educational attainment, measured in certification is, for these reasons, likely to be the strongest common trait amongst intellectuals.

While illiteracy rates (summarized in Figure 1) had dropped to below 5% before the end of World War II (Andorka & Harcsa 1990c), it was clearly the advent of state socialism that marked spectacular increases in formal educational attainment among Hungarians. As we summarize in Figure 2, the high school completion rate was hovering around the 5% mark until the late 1940s, to pick up steeply and rise to the level of 22.3% by 1980 (Andorka & Harcsa 1990c). College or university completion rates establish a very similar pattern: They increased from a three-decade constant of 1.3% until 1949 to 4.9% in 1980 (Andorka & Harcsa 1990c).

We consider the number of books an individual owns to measure physical possession of a cultural commodity. If books are acquired through inheritance, then a double dose of Bourdieu-esque "disposition of the past" is present — some bit of family wealth in a cultural commodity form, some active perception — that is, the choice to keep the books rather than unload them for cash. We could argue that the model depicts something else — income — and that an individual might become wealthy and then buy books. While this might initially appear to be a causal ordering problem, the choice of what to buy remains a disposition of taste, an effect of cultural capital. Individuals can spend money in a number of ways, so if the choice is a cultural commodity, especially of the "low-tech" variety represented by books, skeptics still must explain this type of consumption. Buying and owning books still falls into that category of status which Weber labeled "style of life." Cultural capital strongly influences this process; parental wealth/income only weakly. Current income is only a force offering some new set of chances to put cultural capital into action. To the extent that the number of books is an inherited pattern — both in the sense of inheriting the volumes and a specific system — the number of books an individual owns can be considered a marker of the *Bildungsbürger* ethos.

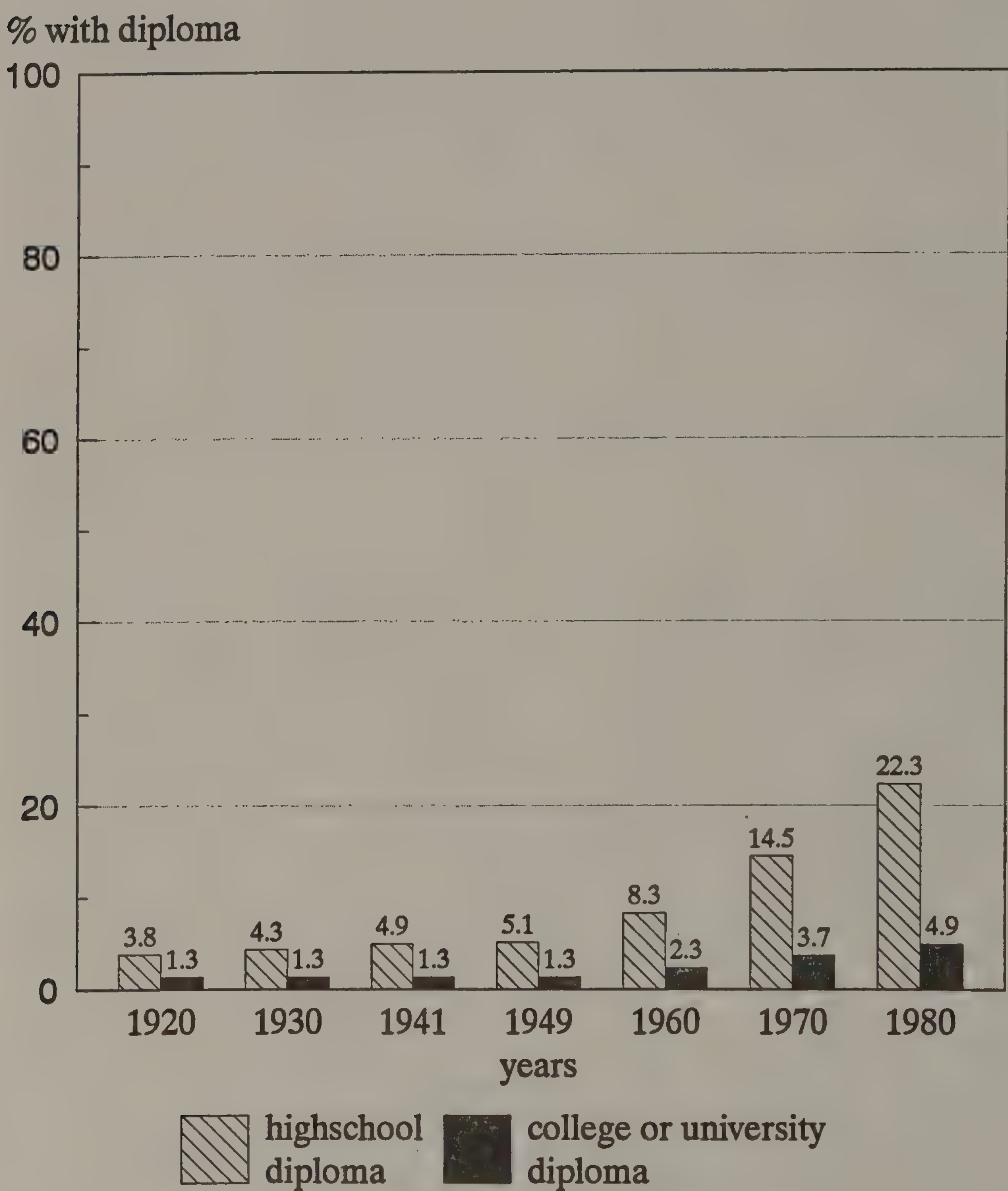
Number of books is distinct from certification in that it does not appear to have the same direct, intentional effect on income — i.e., it is difficult to envision a short-term individual class-mobility strategy based on the act of buying books — and that it offers a specific possibility for the functioning of inherited cultural commodities. A relatively greater number of books can be, however, an integral part of a long-term personal strategy of intellectual "catching-up": Individuals who came from relatively disadvantaged backgrounds, and who engaged in upward mobility on the educational ladder, sometimes increased their spending on cultural commodities. This was especially possible under the state socialist system of cultural subsidies that

FIGURE 1: Illiteracy Rates, Hungary, 1880-1970[a]



[a] Source: Andorka et al. 1990

FIGURE 2: Percentage of Population with Secondary and/or Tertiary Education, Hungary, 1920-1980[a]



[a] Source: Andorka et al. 1990

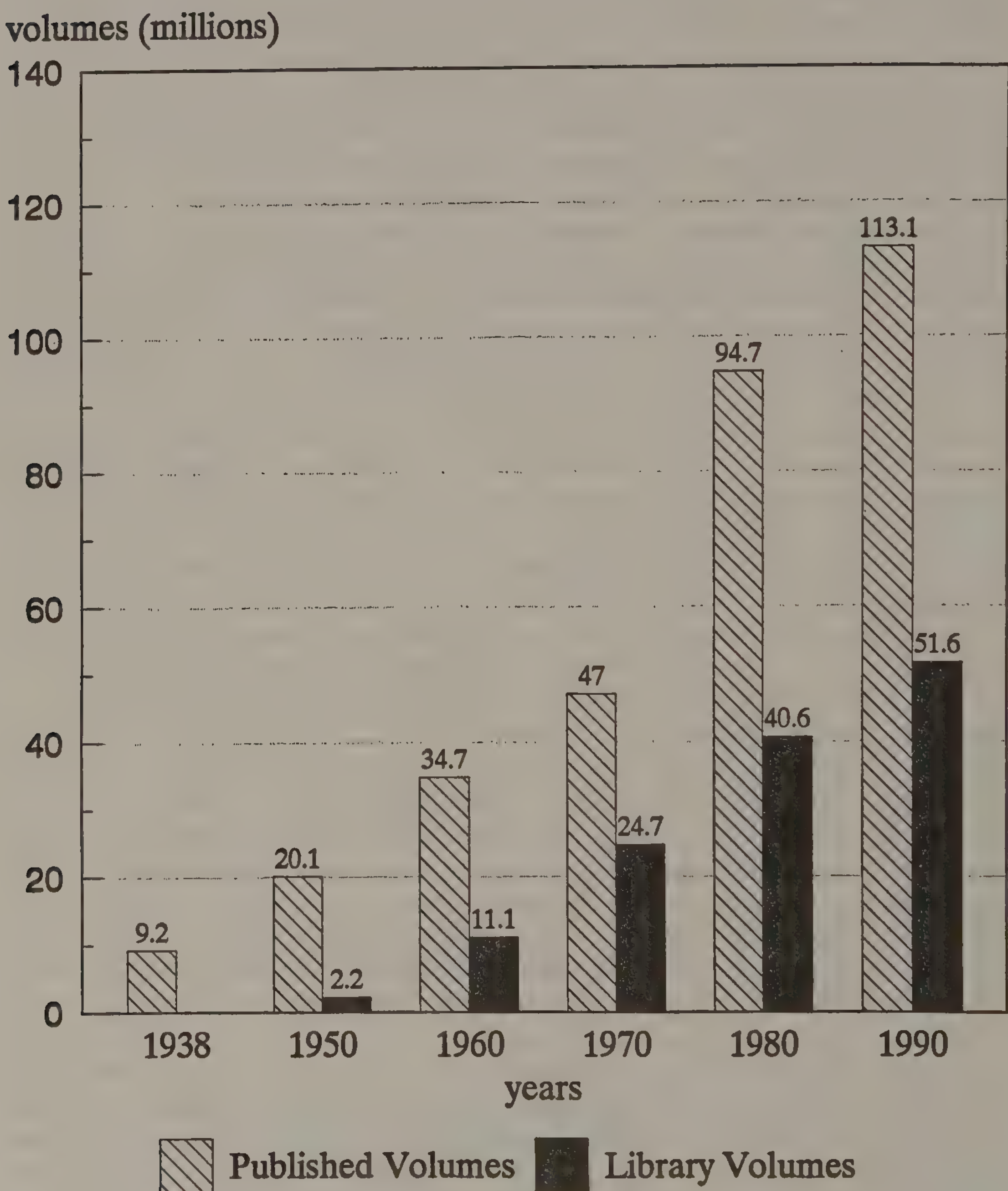
made book purchases very accessible. The number of books published in Hungary increased from 2,438 in the last prewar year to around 8,000 per year by the 1980s (KSH 1991:19). With illiteracy rates reduced by the time of the onset of state socialism, and spectacular increases in the total output of book publishing houses (Figure 3) rising from 9.2 million volumes in 1938 to 113.1 million in 1990 (KSH 1991:19), Hungary can be considered a highly developed and widely accessible book market by the time of the survey. Such catching-up strategies suggest the possibility of the workings of an acquired *Bildungsbürger* pattern among the *nouveaux intellectuels* taking advantage of the socialist state's efforts in increasing cultural capital among the population in general.

Because proficiency in foreign languages comes in large part from a combination of schooling, parental input, and travel experience abroad, language ability is a gauge of the educational component of cultural capital. It also measures education independent of formal institutional recognition. Bi- or multilingual parents are, of course, in a strong position to establish this skill as heritable education, just as wealthy or well-connected parents can provide opportunities of travel. Growing up abroad — as is the case with children of elite intellectuals, foreign traders, and members of the diplomatic service — virtually guarantees advantages in this area. Foreign language proficiency, however, is not necessarily connected to educational attainment or privileged background: Members of Hungary's ethnic minorities, ethnic Magyars who grew up in the neighboring countries as ethnic minorities there and subsequently immigrated to Hungary, documented and undocumented former guest workers abroad (with experience ranging from East Germany, the Soviet Union or Libya to West Germany and other west European countries), along with a small number of foreigners who married Hungarians and settled in Hungary — all these people are found in the group marked by high levels of foreign language proficiency.[14]

Book reading is a gauge of the behavioral component of habitus. We can assume that this measure is largely independent of income or parental income as free libraries and book lending between individuals allow people to read regardless of whether they own the books. Public libraries that were practically free showed impressive growth during the four decades of state socialism in Hungary. As Figure 3 shows, their total holdings grew almost 23 times between 1950 and 1990 (KSH 1991:19). If one appreciated reading, it was certainly there for the taking, especially in the urban setting.

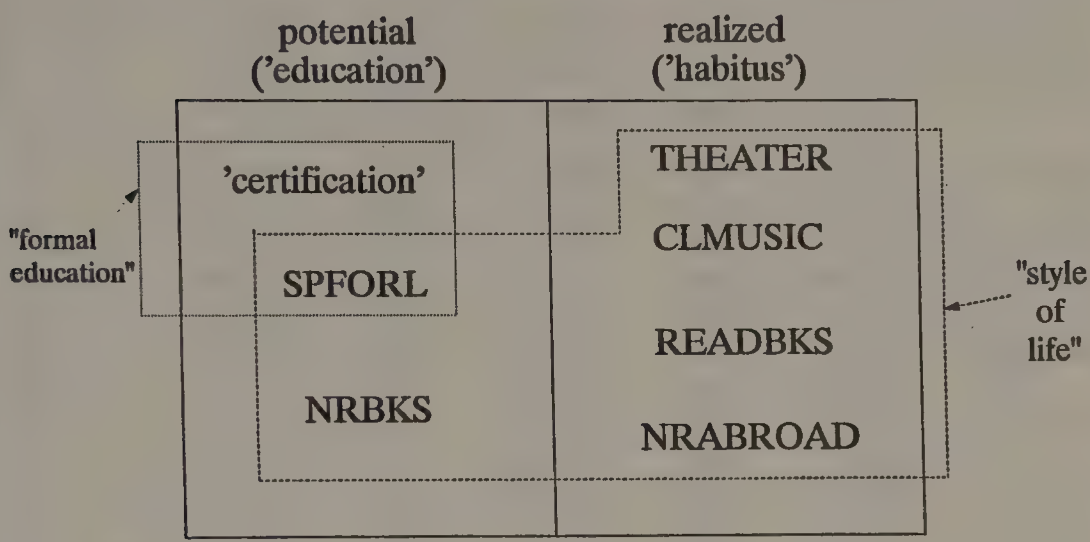
Theater and classical music measure yet different pieces of the habitus in cultural capital. These are markers of direct cultural consumption without producing tangible and displayable commodities. Perhaps they reflect some weak external influence from income in that those who can afford tickets without hardship may attend performances more often. But here again, the choice of attending artistic performances over other alternatives is the important moment. State subsidies also reinforce the accessibility of these cultural commodities. Before the collapse of the socialist system of cultural management, all Hungarian public cultural institutions, such as theaters and concert halls, received substantial state subsidies, some exceeding the equivalent of several times their ticket revenue. In 1987, for instance, the 41 Hungarian full-time professional theaters gave 13,082 performances (KSH 1991:289) so that they saw

FIGURE 3: Annual Volume of Books Published (1938-90) and Library Holdings (1950-90), Hungary (millions)[a]



[a] Source: KSH, 1991

FIGURE 4: Conceptual Model and Variables of Cultural Capital



an average of 56 theater visits for every 100 members of the population (Andorka & Harcsa 1990b:211-12).[15] Attendance at classical music concerts grew from 2 to 10 per 100 population between 1960 and 1990 (Andorka & Harcsa 1990b:211-12). The mean ticket price for professional live theater performances was HUF 59.70 — the equivalent of U.S.$1.27 — in 1987 (KSH 1991), less than one percent of the average net monthly income recorded in official statistics.[16] State expenditures on "collective consumption in kind" can be seen as cultural subsidies to both the process of artistic production and the consumer — the latter being an instance of the positive feedback effect between intellectualness and access to redistributive advantages from the state.

Travel is yet another variable in the nexus between income and cultural capital Weber called "style of life." Without habitus and income, individuals will not travel, but, as with books, classical music, or theater, travel cannot be understood as simply a function of income. Travel is a reproduced behavior and something that one's parentage undoubtedly helps individuals appreciate, decode, and consume. This measure may also suggest an undrawn link between occupational prestige and travel as an element in status: Travel could be part of the prestige of some job that might, in turn, be seen to feed income.[17]

This question emerges here: Does this amalgam of measures adequately capture "intellectualness"? Specifically, how does this particular set of measures of cultural capital add explanatory power to a model beyond the basic stratification variables in the context of Hungary in 1986? Decomposing cultural capital into these variables allows analysis impossible at higher levels of abstraction. Many studies of cultural capital treat cultural consumption without a theoretical justification for what is consumed (Nan Dirk DeGraaf 1988; P. DeGraaf 1991;

TABLE 2: Hierarchial Regression Effects of Cultural Capital on Log (Income), Hungary, 1986-87 for Those Who Have Income, OLS Metric Coefficients[a]

	Baseline	Certi-fication	Number of Books	Foreign Language	Read Books
Constant	3.514***	3.373***	3.512***	3.477***	3.435***
Age	.001***	.002***	.001***	.001***	.002***
	(.081)	(.112)	(.082)	(.085)	(.105)
Gender	-.163***	-.165***	-.165***	-.164***	-.165***
	(-.321)	(-.323)	(-.324)	(-.322)	(-.323)
High party	.133***	.085***	.124***	.127***	.122***
	(.115)	(.074)	(.107)	(.109)	(.105)
Low party	.129***	.096***	.124***	.126***	.121***
	(.146)	(.109)	(.140)	(.142)	(.136)
Work	.205***	.198***	.201***	.207***	.205***
	(.371)	(.359)	(.364)	(.374)	(.370)
Both sectors	.115***	.123***	.115***	.117***	.116***
	(.158)	(.169)	(.157)	(.161)	(.159)
Nonstate	-.008	-.012	-.016	-.013	-.012
	(-.005)	(-.008)	(-.010)	(-.008)	(-.007)
Budapest	.178***	.134***	.159***	.167***	.160***
	(.285)	(.215)	(.256)	(.268)	(.257)
Other city	.059***	.043***	.054***	.057***	.050***
	(.114)	(.082)	(.103)	(.108)	(.096)
Basic education		.107			
		(.201)			
High school		.189**			
		(.317)			
College		.254***			
		(.235)			
University		.295***			
		(.251)			
Number of books * 10^3			.038***		
			(.098)		
Foreign language				.026***	
				(.071)	
Read books					.032***
					(.093)
Theater					
Classical music					
Trips abroad					
Adjusted R^2	.4038	.4443	.4122	.4083	.4107
R^2_{cha}		.0413***	.0086***	.0048***	.0072***

[a] β's are in parentheses.

TABLE 2: Hierarchial Regression Effects of Cultural Capital on Log (Income), Hungary, 1986-87 for Those Who Have Income, OLS Metric Coefficients[a] (Continued)

	Theater	Classical Music	Trips Abroad	Education	Habitus	Cultural Capital
Constant	3.509***	3.515***	3.509***	3.384***	3.452***	3.382***
Age	.001***	.001***	.001***	.002***	.002***	.002***
	(.089)	(.080)	(.078)	(.110)	(.101)	(.113)
Gender	-.167***	-.164***	-.159***	-.165***	-.163***	-.164***
	(-.328)	(-.323)	(-.313)	(-.324)	(-.321)	(-.322)
High party	.123***	.131***	.124***	.084***	.110***	.080***
	(.107)	(.113)	(.107)	(.073)	(.095)	(.069)
Low party	.125***	.129***	.121***	.095***	.114***	.092***
	(.141)	(.146)	(.136)	(.108)	(.128)	(.103)
Work	.200***	.204***	.201***	.196***	.198***	.192***
	(.361)	(.369)	(.364)	(.355)	(.357)	(.348)
Both economies	.117***	.115***	.118***	.122***	.119***	.124***
	(.161)	(.158)	(.161)	(.168)	(.164)	(.169)
Nonstate	-.015	-.010	-.020	-.014	-.026	-.023
	(-.009)	(-.006)	(-.013)	(-.009)	(-.017)	(-.014)
Budapest	.157***	.172***	.159***	.130***	.134***	.115***
	(.252)	(.276)	(.255)	(.208)	(.214)	(.184)
Other city	.055***	.058***	.052***	.041***	.043***	.036***
	(.106)	(.110)	(.100)	(.079)	(.083)	(.069)
Basic education				.107		.099
				(.201)		(.185)
High school				.188**		.172**
				(.316)		(.288)
College				.253***		.227***
				(.234)		(.211)
University				.292***		.264***
				(.249)		(.225)
Number of books * 10^3				.016**		.011
				(.041)		(.028)
Foreign language				-.008		-.014*
				(-.022)		(-.038)
Read books					.023***	.007
					(.066)	(.022)
Theater	.007***				.005***	.003**
	(.105)				(.075)	(.046)
Classical music		.006***			.002	.000
		(.052)			(.015)	(.002)
Trips abroad			.003***		.003***	.002***
			(.119)		(.097)	(.072)
Adjusted R^2	.4132	.4062	.4165	.4456	.4262	.4520
R^2_{cha}	.0097***	.0026***	.0129***	.0430***	.0232***	.0502***

Mean log (income) = 3.723
Standard deviation = .254
(n = 2,518)

* $p < .05$ ** $p < .01$ *** $p < .001$

Ganzeboom, DeGraaf & Robert 1990; Katsillis & Rubinson 1990; Robert 1991). In their work on the intergenerational reproduction of cultural capital, Ganzeboom, DeGraaf and Robert (1990), for instance, test variables representing cultural capital without specifying any conceptual relationship among them. Further, their test is not contextualized in terms of any class or group receiving the benefits of the resulting stratification. The Konrád-Szelényi thesis allows us to see beyond the mere existence of some source of inequality and places the results in relation to the socioeconomic power structure.

Figure 4 places the variables selected for the analysis in the context of our conceptual model. This set of variables is broken down into Bourdieu's two basic types, education and habitus. We have superimposed Weber's categorization of style of life versus formal education on Bourdieu's framework to show the conceptual overlap between them.

For our purposes, we consider people scoring high on these measures to be exhibiting "intellectualness," but it may be possible to score high on some of the less interrelated variables and not be an intellectual. For example, a retired person could read many books without being adequately characterized by any of the other internal measures of cultural capital; it is one of only a few forms of entertainment open to a less-than-mobile pensioner. But someone scoring high on two or more measures, and especially from both categories, suggests the presence of cultural capital.

The statistical analytical method chosen for this analysis is hierarchical regression analysis. First we establish a baseline model predicting the logarithm of income. Then we test the influence of the individual elements of cultural capital in a series of models of income, examining the significance of the improvement of the model for each of the component variables vis-à-vis the baseline model.

Results

In Table 2, each column is a separate hierarchical model showing the effects of a specific component or concept within cultural capital. The far left column, "baseline," models only the stratification variables upon which the cultural capital models improve. Hierarchical regression models provide a means to test blocks of variables in their conceptual whole against the effects of individual measures of cultural capital.

As Table 2 reveals, all the baseline variables except one are significant ($p < .001$) predictors of income in the context of the baseline model. Net of the effects of withdrawing from the labor force, each year that a respondent grows older (*age*) yields higher incomes.[18] Furthermore, gender effects are pronounced and women's income disadvantage is very resistant to variation throughout all models. The party membership variables both contribute to higher total incomes, with the advantage of *high party* slightly larger than *low party*.

Among the baseline variables that describe work, labor force participation has the largest effect on income. Working in both the state and nonstate sectors proves to be a more effective income-earning strategy than only holding a state-sector job (the reference category). The radical withdrawal from the state sector

into exclusively private sector work was not a path to higher incomes. In 1986, this was a very small group; 79 cases in our sample.[19]

The urban/rural contrast and urban primacy also effect incomes. Those who lived in Budapest attained a 17.8% higher income; those who lived in other large cities, 5.9% when compared to the remainder of the population. These results suggest that there were economic opportunities and higher wages in the urban areas and the capital. Such differences also affect cultural consumption as they coincide with cultural institutions such as theater, ballet, and musical groups that may not be seen in the outlying areas. In all, the baseline model suggests a diversified system of income variation. With 40.38% of the variance in income explained, it provides a strong baseline for the examination of the marginal effects of various components of cultural capital on income.

Next, the marginal effects of each cultural capital variable or set of variables is tested beyond the baseline model. The first set of variables pertaining to certification was hypothesized earlier to be the most important component of education and of cultural capital. The increment explained variance associated with certification — $R^2_{cha} = .0413$ — lends strong support to that expectation.[20] We find a systematic pattern associated with the income effects of each level of certification. Higher levels of certification are associated with higher levels of remuneration. The metric coefficients indicate a monotonically increasing pattern as we move from "incomplete elementary school" to "university diploma." As we move from lower to higher ranks of certification, we can take it with ever greater certainty that the statistical effects of certification on income are not attributable to chance.

Certification eliminates one-third to one-half of the explanatory power of party membership or rank. The fact that no other measure of cultural capital tested in these models deducts from the income-generating statistical effects of party membership and rank to the same degree as certification suggests that those with formal education are drawn into membership and the more active party ranks. The partial overlap between party rank and certification levels is also an implicit empirical measure of the degree to which the merger between the political and the intellectual elites took place in Hungary, as predicted by Konrád and Szelényi ([1974] 1979) over ten years before the time of our survey. These results suggest a *partial* overlap: Although the inclusion of the certification variables reduces the magnitude of party membership and rank, the direction of the effect of the measures of party membership and rank remain the same. This suggests that certification does not simply determine party membership and rank.[21] By the late socialist period, the merger of the party and certification hierarchies remained rather limited.[22]

The certification variables affect party rank unequally such that *high party* contributes slightly less to income than *low party*. In this we see that not only are a significant portion of the benefits of party membership from educational credentials, but that — net of education — it is joining the party, not working to become a high-ranking party official, that increases income the most. This result indicates that the system of rewards advantaged those with a state-sector job (apart from political involvement) *and* a low, often nominal, level of political commitment to the vaguely defined objectives of the ruling party. This combination earned higher incomes than those with a higher-level political

obligation and the responsibility that came with choosing a more exclusively political career.

An additional, conceptually important result pertains to the changes in the strength of the *age* coefficient. In the context of the baseline model *age* is significant ($p < .001$) with a weight of $\beta = .081$; however, when certification is added, the weight of age increases substantially ($\beta = .112$). This indicates that each one-year change in age is associated with a larger increase in income. In substantive terms, the result implies that the expected gradual increases associated with age (having to do with job seniority, experience, etc.) can be observed more within each level of certification: Educational degrees demarcate from each other somewhat separate systems of seniority and experience.

Finally, certification has modest effects on labor force participation and the sector of employment. The decline of the *work* coefficient suggests that more of the educated tend to work. The increase in the *both sectors* coefficient suggests that those working in the state and nonstate sectors have fewer formal educational credentials than those solely within the state sector. This suggests that, by 1986, informal income earning practices had penetrated all segments of the Hungarian society, quite irrespective of educational attainment. Second economy participation was not, then, a unique economic outlet for the intelligentsia.

The geographic variables are substantially reduced with the control for education. This shows the concentration of those with higher education in the urban areas. Again, this reenforces the importance of controlling for settlement type in examining cultural practices associated with intellectuals.

When examined separately from each other, all six of the remaining variables of cultural capital show significant positive effects on income, although the magnitude of their coefficients is more modest than those for certification. In contrast to certification, the inclusion of the remaining measures of cultural capital does not take away from the relative marginal importance of political capital: The coefficients of party membership and rank remain virtually constant throughout the six alternative models. Possession of books, ability to speak a foreign language, book reading, theater and concert attendance, and travel abroad are characteristics and behaviors that are not, by and large, associated with the degree of affiliation with the ruling party.[23] The returns to cultural capital also do not vary by economic sector.

We are not attempting to measure "party culture," although such a phenomenon did exist. The party had special schools, encouraged specific kinds of literary and art expression, arranged its own social retreats and, as in the case of high cadres, was a vehicle for organizing hunting parties and special resorts. There was some overlap between certain elements of party culture and cultural capital (high-status culture) since ruling groups typically adopt some of the practices of their predecessors. However, the efficacy of cultural capital as we use it is in its independence from direct ruling class control; it is a marker, largely produced through educational institutions and the family, that allows people to discern a more complex social reality. Cultural capital is thus not simply the sum total of the patterns of conspicuous consumption by the ruling

class or the genres of artistic expression approved of by the culture and propaganda departments of the Central Committee.

Table 2 (the three columns on the right-hand side) shows the test of the relative magnitude of these measures of cultural capital in the context of their various combinations. The "education" column estimates the combined effects of the three "education" components of Bourdieu's cultural capital — "certification," the number of books and foreign language proficiency — on earnings. The modest independent effects of *number of books* and *foreign language* are overshadowed by the power of certification as the strongest predictor of income: The effect of the number of books is cut by over half. The influence of fluency in a foreign language diminishes completely. In contrast, reductions in magnitude and statistical significance in each of the measures of certification are negligible.

The next column estimates the joint effects of our variables pertaining to the habitus component of cultural capital. Three of the four variables that are significant in the separate models — *read books, theater,* and *trips abroad* — maintain much of their effects. The collective effect of the habitus variables ($R^2_{cha} = .0232$) is noticeably smaller than that of "education" variables ($R^2_{cha} = .0430$), although both are statistically significant ($p < .001$).

The panel at the extreme right-hand side of the table presents our full model of income variation. It suggests that certification is the most important component of our model of cultural capital: All measures of certification preserve their direction and much of their magnitude while greatly reducing the effects of the cultural capital variables. Reading and possessing books and listening to classical music become insignificant in the face of certification and the other cultural controls. *Theater* and *trips abroad* retain a modicum of their explanatory power and ability to distinguish intellectuals net of education, location, and the other controls.

In the context of the full model, the effect of foreign language proficiency becomes negative and statistically significant at the $p < .05$ level. We attribute this change to the following mechanism: The inclusion of a large number of cultural capital variables — including measures for both formal certification and habitus — captures the effects of formal education as well as the *Bildungsbürger* patterns. As a consequence, the remaining effects of proficiency in a foreign language are those that have to do with ethnic minority status or personal background as a former member of the large ethnic Hungarian minorities of the neighboring states.[24] We can discard possible confounding effects between *foreign language* and age, labor force participation and involvement in bisectoral income-earning activities.[25]

That the full model demonstrates an effect for certification and some of the cultural capital variables is a substantial finding. By their very nature, cultural distinctions are more difficult to recognize, quantify, and measure than ascribed characteristics or the locations of home and work.

Discussion and Conclusion

The picture that emerges from this study is one of a material force embodied in the deployment of cultural capital. Rather than an exclusive view of culture as language and shared understanding of ideas, as an art that is to be passively appreciated in quiet separation from those mundane aspects of life that have to do with social action and orientation toward economic goals, in these results we see culture taking on a commodity-like form. The more of it someone has, the more can be converted into systematic distributional advantages or used for the satisfaction of economic ends. Our results can be summarized as follows:

First, the field of cultural capital has a hierarchical internal organization, with the conferral of certification as its focal component. It is mainly through certification that cultural capital was transformed into systematic distributional advantages in the Hungary of the late state socialist period.

Second, levels of certification operate as markers allowing the operation of finer distinctions, such as age, mainly within particular levels of certification.

Third, the strength of the relationship between level of certification and income increases as we pass from the lower to the higher levels of formal education, suggesting that the closure of the ranks of certification increases toward the higher end of the certification scale.

If one accepts, as we initially did, the Konrád-Szelényi thesis concerning the class nature of state socialism with intellectuals as its privileged beneficiaries, then the results presented above indicate the existence of basically two main routes to class mobility under late state socialism: (1) via low-level affiliation with the ruling party and (2) through increasing one's potential cultural capital, specifically, the components that correspond to Weber's "formal education." We have found support for a broader Bourdieu-esque conceptualization of the field of cultural capital in the significance of the single measures of "style of life" and specifically, habitus variables. The statistical models do suggest to us, at the same time, that the coherence of cultural capital is organized by the second Weberian component of status, formal education, as distinct from the style of life.

The implications of this set of findings are two-fold. First, it portrays late state socialism as a credentialing society whose system of basic material rewards can be apprehended through a Weberian-Bourdieu-esque model of formal educational achievement. Second, the demonstrated — more moderate but also somewhat resilient — positive effects of habitus variables suggest that a human capital interpretation of our findings is inadequate unless contending explanations are able to draw an unambiguous causal link between the frequency of such components of habitus as book reading, theater attendance, and travel abroad as individually utility-maximizing income earning strategies.

If it is not cultural capital in general, but a specific set of cultural capital variables that confer distributional advantages to individuals in a state socialist society, then there is reason to argue that the main avenue for the conversion of cultural capital to income leads through labor market processes, specifically through the standardized allocation of segregated job opportunities according to levels of formal education. As the overwhelming majority of employment under late state socialism continued to be in the state sector, our results

reinforce the image of an educational merit-based system of state-allocated inequalities transferred primarily through labor market position.

Our results show that the inequalities of the pre-state-socialist order are in part reproduced and in part reorganized. As the increases in various measures of education and habitus attest, it would be wrong to assume that the rich or the old bourgeoisie possess all the cultural capital via inheritance. Still, despite the greater access to education and culture provided in the Hungary of the 1950s through the 1980s, it would also be mistaken to argue that anyone could wield cultural capital as a mechanism for greater income. Not everyone will know that such a route to privilege exists, nor would they all succeed in the endeavor. Just as the majority in a capitalist society cannot pull on their bootstraps to change their class position, individuals in a state socialist system cannot use cultural capital as a universal access code for economic power.

These findings allow a further speculation concerning a possible interpretation of state socialist practices in cultural capital. If certification is so much more important than all other components of cultural capital as it appears from the above results, then we should expect marked differences in the class effects of cultural subsidies dispensed by the socialist state. An implication of our findings is that the potential impact on class mobility of state spending on the equalization of access to education is much greater than that of all other forms of cultural subsidies. Let us note that, according to Andorka and Harcsa (1990b), a marked drop in state funding for education — implying a serious closure in access to certification — occurred during the 1970s, exactly the period in which Konrád and Szelényi documented the appearance of intellectuals on the road to class power.

Our model can be improved in a number of ways. Time series data would enable researchers to identify the exact dynamics of changes in the conversion of cultural capital to income in Hungary. An inclusion of data for the parents' and grandparents' generation would make it possible to separate more clearly the effects of the *Bildungsbürger* ethos from individual cultural mobility. A replication of at least the most crucial questions of the survey we have analyzed would make it possible to design a before/after study to decipher the effects of social change after state socialism on the processes of conversion. The model could be further specified by seeking out associated variables of intellectualness (such as the effect of access to, and the use of, social networks[27]) and by specifying the labor market route of conversion through some measure of occupational position and institutional power.

Notes

1. Between 1962 and 1987, the ratio of the net volume of officially recorded income in the top and bottom income deciles varied between 3.8 and 5.18, estimated at 4.64 in 1987 (Andorka & Harcsa 1990a:109).

2. The concepts come from German historiography (e.g., Kocka 1981, 1987; Lepsius 1987; Rüschemeyer 1987). Due to the historical lateness of German industrialization, the bourgeoisie was fragmented in part due to its ties to the disparate organs of the German state (Blackbourn 1991). The educated and propertied bourgeoisie occupied separate spheres with different relations to the market, sources of economic security, types of education and recruitment to their respective professions. Arguably, it was this split that prevented the German bourgeoisie

from assuming the unifying role its counterparts had fulfilled before in France and England until World War I (Maier 1975). Zoltán Tóth (1991) offers an erudite application of this conceptual apparatus to Hungarian social history.

3. In this analysis, we have no room to address the obvious ambiguity of the English translation of the Weberian German term *ständische Lage* — "status" vs. "estate" or "rank" (Kozyr-Kowalski 1983; Wenger 1980). As shorthand, we adhere to the usage adopted in stratification studies. This problem is taken up in much more detail elsewhere (Böröcz 1995).

4. We are aware of Abel and Cockerham's (1993) recent warning concerning the mistranslation of Weber's original concept of *Lebensführung* as "style of life" instead of the more exact "life conduct."

5. Christian Joppke (1986) uses the intuitively appealing terms "institutionalized" versus "internalized" cultural capital to denote the distinction between education and habitus. We avoid using this terminology because they are conceptually not orthogonal (i.e., "internalized" cultural capital or, habitus, is no less "institutionalized" than education and *vice versa*).

6. In reality, none of the "socialist" states managed to abolish completely the category of non-state property: A non-collectivized peasantry of various proportions remained, and so did numerous private artisans and other "own-account workers." However, the relative importance of this property form has been minuscule due to serious constraints on the process of capital formation by the "socialist" state, in effect (1) limiting the number of employees to be hired by such establishments (in Hungary, that limit had been 4 until the mid-1980s when it was raised to 15); (2) restricting the availability of capital goods to very small-scale and low-profitability equipment; (3) withholding business credits and systematically encouraging (sometimes even subsidizing) conspicuous final consumption; and (4) the systematic harassment of those successful in being able to accumulate notwithstanding the above measures.

7. Konrád and Szelényi summarize their position on the class reproduction mechanism under state "socialism" this way:

No serious claim can be made about the formation of a class unless the analysis can identify a position from which these agents can exercise class power. In the *Intellectuals on the Road to Class Power*, we believe we identified such a position in the institutions of the redistributive economy. In that book, our argument was that in a statist economy, the redistributor occupies a position that is analogous to that of property owners in a capitalist society (Konrád & Szelényi 1991:352-53).

8. The sample is structured according to settlement type, age group and gender. The data come from three subsamples of 1,000 respondents each (TÁRKI 1988). The sample was managed by Éva Iván. Sociologists Róbert Angelusz, Ágnes Bokor, Tamás Kolosi, and Róbert Tardos were the principal investigators of various parts of the survey. Some results have been presented in Angelusz and Tardos (1991).

9. Logarithmic transformation of total income was used to capture the curvilinear nature of income data. This way, the metric regression coefficient can be interpreted as percent increment in income per unit increase in the predictor variable (Petersen 1989:228).

10. Following notation used in international banking, Hungarian Forints are abbreviated as HUF.

11. Income in no way captures the full range of inequalities in Hungary. This is because, first, a high proportion of income is accessed through collective consumption that is not individually divisible. Second, the fluidity of monetary incomes is much higher than some in-kind assets, namely housing. The correlation between total income and estimated housing value in our sample was .118. It was partly this insight that initiated Konrád and Szelényi's theory.

Despite these limitations, the Hungarian economy in 1986 was the product of numerous reforms which monetized many formerly in-kind transactions with new accounting practices. The central bank even held hard currency reserves approximately equal to the value of Forints saved. Moreover, while Konrád and Szelényi drew their hypothesis from housing inequality, they — and much of the literature emerging after their work — treat their proposition as generalizable.

12. It also controls for absence from the formal labor force for other reasons, e.g., student or homemaker status, persons on maternity leave, etc.

13. The categorization of settlements into "urban" (*város*) vs. "rural" (*nagyközség* or *község*) and the various finer gradations within each of those categories (which we do not use here), was an important manifestation of the manifold classificatory struggles that took place during the last three decades of state socialism in Hungary. It reflects complex considerations pertaining to area and population, infrastructural endowments, administrative function, the size and character of the enterprises, and the intricate issue of the "significance" of all of the above as filtered through the capillaries of the political system.

14. For more on some of these flows, see Böröcz 1992.

15. Theater statistics also reveal sharp geographical inequalities: Theater visits per 1000 population ranged between 1479 in the capital city, Budapest, and as low as 46 and 81 in some of the most rural counties (KSH 1991:287).

16. Theater attendance peaked sometime around the early 1960s with almost 6.5 million visits in this country of 10 million inhabitants. The number of theater visits has been on the decline since then, falling slightly below 5 million by 1990 (KSH 1991:19).

17. Our data allow only a blurry measure of travel abroad: the number of trips. This admittedly confounds the purpose of the trip (business or pleasure), the destination (East Bloc or the West), and distance (neighboring country vs. others). While foreign travel to the West required exit and entry visas and hard currency, by 1986 the majority of administrative travel restrictions had been eased and visas were rarely denied in Hungary. Hard currency was quite widely available through the second economy and, specifically through contact with incoming Western tourism (Böröcz 1996). So while *trips abroad* must be viewed with caution, in the Hungarian context travel is closer to a cultural choice than exclusively the province of state bureaucrats and those with party connections.

18. While the weight of the *age* variable might appear small compared to *work*, for example, the effects of the other variables are built up over the life course and might be more accurately compared to the effects of 10-20 year periods (although that is not our purpose here).

19. The disadvantage associated with being only in the private sector is a strong finding that contradicts much of the existing literature about the dynamism of entrepreneurship and the power of the free-market sector over the state sector. At least in 1986, that logic does not appear to hold; second economy production is rather an important sideline of work, not an alternative. Space does not permit us to explore this further here.

20. In an alternative model not presented here, we represented certification in a single rank-order variable, assuming equal distance among each of its categories. That step also yielded an increment significant at the $p < .001$ level.

21. The low coefficients in the zero-order correlation matrix between party and certification variables (see below) suggest a similar conclusion.

	Basic Education	High School	College	University
High party	-.160	.049	.103	.148
Low party	-.139	.054	.109	.082

22. This conclusion is reinforced by an alternative model in which we tested for the additional effects of an interaction term composed by (1) a dummy variable separating all party members from nonmembers, and (2) one grouping those with tertiary education apart from those less educated: The *F* test of the improvement yielded by the model with the interaction term over our full model was not significant statistically.

While the party and educational hierarchy were not identical, educational credentials were one mechanism the party used to increase its legitimacy. From the late-1960s onward, the party claimed that technical competence, not political loyalty was its first consideration in promotion of industrial managers. During this period, the party also maintained its own schools and special avenues of admission to the more prestigious public universities and institutes (to permit its cadres to attend despite strongly meritocratic admissions policies).

23. The zero-order correlation coefficients between party membership/rank and the remaining measures of cultural capital signal association that is sporadic if not absent:

	Number of Books	Foreign Language	Read Books	Theater	Classical Music	Trips Abroad
High party	.097	.088	.118	.101	.052	.088
Low party	.064	.048	.091	.036	-.004	.090

24. Ethnicity- and immigration-related variables are unfortunately not available in the dataset.

25. The zero-order correlations between these variables are, respectively, -.021, -.012 and -.063.

26. See, e.g., Angelusz and Tardos 1991; Böröcz and Southworth 1995.

References

Abel, Thomas and William C. Cockerham. 1993. "Lifestyle or Lebensführung? Critical Remarks on the Mistranslation of Weber's 'Class, Status, Party'. *The Sociological Quarterly* 34:551-56.

Andorka, Rudolf, and István Harcsa. 1990a. "A lakosság jövedelme." Pp. 97-117. in *Társadalmi riport*, edited by Rudolf Andorka, Tamás Kolosi, and György Vukovich. Budapest: TÁRKI.

____. 1990b. "Művelődés." Pp. 208-14 in *Társadalmi riport*, edited by Rudolf Andorka, Tamás Kolosi, and György Vukovich. Budapest: TÁRKI.

____. 1990c. "Oktatás." Pp. 40-56 in *Társadalmi riport*, edited by Rudolf Andorka, Tamás Kolosi and György Vukovich. Budapest: TÁRKI.

Angelusz, Róbert, and Róbert Tardos. 1991. *Hálózatok, stílusok, struktúrák*. Budapest: ELTE Institute for Sociology.

Blackbourn, David. 1991. "The German Bourgeoisie." Pp. 1-45 in *The German Bourgeoisie*, edited by David Blackbourn and Richard J. Evans. Routledge.

Böröcz, József. 1992b. "Dual Dependency and the Informalization of External Linkages: The Case of Hungary." *Research in Social Movements, Conflicts and Change* 14:189-209.

____. 1993. "Informality and the Second Economy in East-Central Europe." Pp. 215-44. in *Work Without Protections: Case Studies of the Informal Sector in Developing Countries*, edited by Gregory K. Schoepfle and Jorge Pérez-Lopez. U.S. Department of Labor, Bureau of International Labor Affairs, Washington, D.C.

____. 1995. "Social Change by Fusion?" Paper presented at the conference of the International Institute of Sociology, Trieste, 3 July.

____. 1996. *Leisure Migration: A Sociological Study of Tourism*. Pergamon Press. In press.

Böröcz, József, and Caleb Southworth. 1995. "Kapcsolatok és jövedelem: Magyarország, 1986-87." *Szociòlgiai szemle* 2:25-48.

Bourdieu, Pierre. 1977. *Outline of a Theory of Practice*. Cambridge University Press.

____. 1984. *Distinction*. Harvard University Press.

____. 1986. "Forms of Capital." Pp. 241-60 in *Handbook of Theory and Research for the Sociology of Education*, edited by John G. Richardson. Greenwood Press.

DeGraaf, Nan Dirk. 1991. "Distinction by Consumption in Czechoslovakia, Hungary, and the Netherlands." *European Sociological Review* 7:267-90.

DeGraaf, Paul M. 1988. "Parents' Financial and Cultural Resources, Grades and Transition to Secondary School in the Federal Republic of Germany." *European Sociological Review* 4:209-21.

Djilas, Milovan. 1962a. *Conversations with Stalin*. Translated by Michael B. Petrovich. Harcourt Brace Jovanovich.

____. [1957] 1962b. *The New Class*. Praeger.

Ganzeboom, Harry B.G., Paul M. DeGraaf, and Peter Robert. 1990. "Cultural Reproduction Theory on Socialist Ground: Intergenerational Transmission of Inequalities in Hungary." *Research in Social Stratification and Mobility* 9:79-104.

Gouldner, Alvin. 1979. *The Future of Intellectuals and the Rise of the New Class*. Seabury Press.

Joppke, Christian. 1986. "The Cultural Dimensions of Class Formation and Class Struggle: On the Social Theory of Pierre Bourdieu." *Berkeley Journal of Sociology* 31:53-78.

Katsillis, John, and Richard Rubinson. 1990. "Cultural Capital, Student Achievement, and Educational Reproduction: The Case of Greece." *American Sociological Review* 55:270-79.

Kocka, Jürgen. 1981. "Capitalism and Bureaucracy in German Industrialization before 1914." *The Economic History Review*. 2d series, 33:453-68.

____. 1987. "Bürgertum und Bürgerlichkeit as Probleme der deutschen Geschichte vom späten 18. Jahrhundert." Pp. 21-63 *Bürger und Bürgerlichkeit im 19. Jahrhundert*, edited by Jürgen Kocka. Göttingen: Vandenhoeck & Ruprecht.

Konrád, George, and Iván Szelényi. 1991. "Intellectuals and Domination in Post-Communist Societies." Pp. 337-61 in *Social Theory for a Changing Society*, edited by Pierre Bourdieu and James S. Coleman. Westview Press.

____. [1974] 1979. *The Intellectuals on the Road to Class Power*, translated by Andrew Arato. Harcourt Brace Jovanovich.

Kozyr-Kowalski, Stanisław. 1983. "Max Weber's Theories of Social Estates." *The Polish Sociological Bulletin* 1-4:85-102.

KSH. 1991. *Magyar Statisztikai Évkönyv 1990*. Budapest: Központi Statisztikai Hivatal.

Lamont, Michèle, and Annette Lareau. 1988. "Cultural Capital: Allusions, Gaps and Glissandos in Recent Theoretical Developments." *Sociological Theory* 6:153-68.

Lepsius, M. Rainer. 1987. "Zur Soziologie des Bürgertums und der Bürgerlichkeit." Pp. 79-100 in *Bürger und Bürgerlichkeit im 19. Jahrhundert*, edited by Jürgen Kocka Göttingen. Vandenhoeck & Ruprecht.

Maier, Charles S. 1975. *Recasting Bourgeois Europe. Stabilization in France, Germany, and Italy in the Decade after World War I*. Princeton University Press.

Petersen, Trond. 1989. "The Earnings Function in Sociological Studies of Earnings Inequality: Functional Form and Hours Worked." *Research in Social Stratification and Mobility* 8:220-50.

Robert, Peter. 1991. "Educational Transition in Hungary from the Post-War Period to the End of the 1980s." *European Sociological Review* 7:213-36.

Róna-Tas, Ákos. 1990. *The Second Economy in Hungary: The Social Origins of the End of State Socialism*. Ph.D. diss., Sociology, University of Michigan, Ann Arbor.

Rüschemeyer, Dietrich. 1987. "Bourgeoisie, Staat und Bildungsbürgertum. Idealtypische Modelle für die vergleichende Erforschung von Bürgertum und Bürgerlichkeit." Pp. 101-20 in *Bürger und Bürgerlichkeit im 19. Jahrhundert*, edited by Jürgen Kocka Göttingen. Vandenhoeck & Ruprecht.

Stites, Richard. 1989. *Revolutionary Dreams: Utopian Vision and Experimental Life in the Russian Revolution*. Oxford University Press.

Szelényi, Iván. 1982. "The Intelligentsia in the Class Structure of State-Socialist Societies." In *Marxist Inquiries: Studies of Labor, Class, and States*, edited by Michael Burawoy and Theda Skocpol Supplement to the *American Journal of Sociology* 88:287-326.

____. 1986-1987. "The Prospects and Limits of the Eastern European New Class Project." *Politics and Society* 7:103-44.

TÁRKI. 1988. *A TÁRKI III. felvétel dokumentumai. II. kötet (Miliõk)*. Budapest: Társadalomkutatási Informatikai Társulás.

Tóth Zoltán. 1991. "A rendi norma és a 'keresztyén polgárisodás': Társadalomtörténeti esszé." *Századvég* 2-3:75-130.

Trotsky, Leon. [1933] 1972. *The Revolution Betrayed*. Pathfinder.

Weber, Max. [1920] 1978. *Economy and Society*, edited by Günther Roth and Claus Wittich. University of California Press.

Wenger, Morton G. 1980. "The Transmutation of Weber's *Stand* in American Sociology and Its Social Roots." *Current Perspectives in Social Theory* 1:357-78.

Facial Dominance of West Point Cadets as a Predictor of Later Military Rank*

ULRICH MUELLER, *University of Marburg, Germany*
ALLAN MAZUR, *Syracuse University*

Abstract

Facial dominance of West Point cadets, measured from their graduation portraits, is known to be related to cadets' ranks at the military academy, but it has been reported to be unrelated to their ranks in later career (Mazur, Mazur & Keating 1984). With improved methods of data collection and analysis, we show that cadets' facial dominance, while still unrelated to their ranks at midcareer, is related to promotions in late career, 20 or more years after the portraits were taken. These results suggest that the absence of physical features from current models of status attainment is a serious omission.

Influenced by Blau and Duncan (1967), studies of status attainment have focused on family background and education as primary determinants of one's occupational rank. Despite repeated suggestions that a person's physical appearance also has important effects on his or her life chances (Berscheid & Walster 1974; Gillis 1982; Guthrie 1976; Hatfield & Sprecher 1986; Mazur & Robertson 1972; Patzer 1985), physical features of the face and body have been virtually ignored in modern research on status attainment.

Studies relating physical features to personality were popular during the first half of this century (Gowin 1915). Among the best known is the "somatotype" research of Sheldon (1942), whose claimed relationships between body type and temperament were treated skeptically because of biases in his method, though careful replications by others have verified some of the correlations (Child 1950; Cortes & Gatti 1972). Stogdill's (1974) review of numerous early studies, mostly of school-age groups, concludes that leadership (measured in

* *The authors appreciate the cooperation of the men of the West Point class of 1950 and the advice and support of Colonel Morris J. Herbert (Ret.), U.S. Military Academy, Association of Graduates. The Office of Institutional Research (U.S. Military Academy), the U.S. Army Military History Institute, and the Center for Air Force History helped us in obtaining portraits shown in Figure 1. The research was funded by grant Mu609/5-1 from the Deutsche Forschungsgemeinschaft and by a grant from the Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim/Germany to the first author. Direct correspondence to Ulrich Mueller, Institute of Medical Sociology, University of Marburg, Bunsenstrasse 2, 35033 Marburg, Germany; or mueller2@mailer.uni-marburg.de.*

diverse ways) is correlated with height and with a mesomorphic, or athletic, physique.

It is a common observation that certain individuals have "dominant-looking" faces whereas the faces of others are perceived as submissive. American subjects reliably sort facial portraits along a dominance-submissiveness dimension, and these portraits are given similar ratings in a wide variety of cultures around the world (Keating, Mazur & Segall 1981a, 1981b). Anecdotes suggest that dominant appearance is correlated with actual status attainment, especially within the military, as when Atkinson (1981) refers to the "lantern jaw and chiseled features prized in military officers" or the "square-jawed looks . . . considered vital" in a first captain at the U.S. Military Academy at West Point (1989:118). Also consider this description of a fictitious first captain, written by an academy graduate: "He had one of those young Gregory Peck faces, the dark handsome good looks of a born general. It had always seemed there was an unwritten requirement that first captains and other high-ranking cadets be attractive . . . , not just good-looking, but . . . idols. Statues to the American idea of cadet. . . . At 6'1," 185 pounds, a letter man in soccer and lacrosse, he was the ideal first captain. There was a certain awkwardness — intimidation — in his presence" (Truscott 1978:414).

Mazur, Mazur and Keating (1984) tested these anecdotes, asking if West Point cadets with dominant-looking faces, or who were tall or athletic, were promoted higher than classmates with less impressive looks. Cadets from the class of 1950 were rated as facially dominant or submissive by judges who viewed their graduation portraits. Dominant-looking cadets were indeed promoted to significantly higher ranks than submissive-looking cadets in their junior and senior years at West Point. However the investigators found no relationship between facial dominance and rank attainment over the next 20 years of military service. Athletes enjoyed some advantage in promotion throughout their careers, but height was generally unrelated to rank.

The investigators were puzzled that rank correlated with facial dominance while at West Point but not in later career, and they suggested four plausible explanations.

(1) A cadet's rank is the cause rather than the effect of his appearance in the yearbook portrait. Cadets who enjoy the status of officers communicate their position through erect posture and perhaps through facial gestures, which are recorded in the photograph.

(2) As men enter middle age, physical features lose their earlier relevance for hierarchical placement, at least in comparison with nonphysical features.

(3) Facial features remain important throughout the career, and they continue to influence the promotion board, which requires an updated photograph of each candidate. But a man's face changes over the years, so his appearance in a graduation portrait bears little relationship to his appearance in middle age.

(4) Facial appearance is a leadership quality that matters most in primary groups, when those recommending promotion know the candidate personally. That is the case at West Point where the decision is based on evaluations by cadets and tactical officers, and on promotion to the highest ranks, when the small pool of candidates is again known by those in judgment. At

the junior and middle ranks outside West Point, the candidates are usually unknown to promotion boards and decisions are based on extensive officer evaluation reports, career patterns (whether candidates have had the requisite operational and command positions in the right order and within prescribed time limits), and seniority. Senior officers, however, will all have outstanding evaluation reports with little to distinguish between them. Seniority considerations on this level lose their weight.

Another plausible explanation is sampling bias: West Point graduates make up only 5-10% of all army and air force officers. Also, West Point cadets might be selected more for physical (including facial) appearance than officers who received their commission elsewhere. Thus, while facial appearance may be important for promotions on all hierarchy levels, if we study a sample containing West Point cadets only, this factor may be masked because even cadets with, by West Point standards, average facial appearance, might look more dominant than the majority of junior and middle officers who are not in our sample.

We should then observe a correlation reappear between facial dominance and rank among those men in the sample who obtained promotion to general, since most are from West Point: The percentage of West Point graduates among all brigadier and major generals from World War II until the mid-70s was between 45 and 80%, among all lieutenant and full generals was between 50 and 100% (Moore & Trout 1978); in the early 1980s among all generals it was 40% (Atkinson 1989:496). In our sample, 90% of all promotions to brigadier and 65% of all promotions to major general occurred by 1975, 82% of all promotions to lieutenant and 66% of all promotions to full general occurred by 1980.

Now we are able to evaluate these diverse hypotheses. In 1990, members of the class of 1950 were questioned about their military careers, which by then had all been completed. At the time of the survey, there were still 540 survivors of the original 670 graduates, making the class an ideal object for studying careers in an occupational hierarchy. A refined analysis of these new data provides considerable clarification of the promotion process and some correction of the relationship of facial dominance to military rank attainment.

Careers of the West Point Class of 1950

Like all cadets, the class of 1950 entered the military academy as formally undifferentiated plebes, but in their junior year they were ranked as corporals or privates. In senior year about one-quarter of the men became cadet officers, the remainder, sergeants.

At graduation, all men received the same rank, second lieutenant, and entered one of the army's several branches (e.g., infantry, engineers) or the air force, which at this time did not have its own academy. Men almost always stay in their service, and tend to stay in their branch, for the rest of their professional lives. Most young officers earn advanced degrees and seek a variety of assignments in military schools, command, and staff positions.

Within three years, 8% of the class members were dead (two-thirds of them killed in the Korean War, which began only weeks after graduation). By 1960,

20% of the class had resigned for various attractions outside — and frustrations inside — the army (Butler 1971a). Nearly all those who remained in the military through the 1950s would stay for 20 years (or more) in order to retire with benefits.

While promotion of young officers is nearly automatic through the rank of captain, being determined primarily by amount of time served, early promotion to the rank of major or lieutenant colonel is regarded as an indication of special merit. Of the class of 1950, 6% made lieutenant colonel by 1964, and these men were more likely than their classmates to make general. In fact, speed of promotion to each rank predicts the likelihood and speed of promotion to all subsequent ranks. The sorting of men begins early, separating those who will reach the top from those who will not, as occurs in many other occupational hierarchies (Mazur, Mazur & Keating 1984).

Between 1966 and 1972, about one-third of the men still on active duty were selected to attend a war college, nearly always after promotion to lieutenant colonel. For war college graduates promotion to colonel was virtually certain. Although not a formal requirement, it seems that no one has been selected for general since World War II who has not graduated from a war college. In army jargon, war college is an essential "ticket punch" on the route to general.

Many of the classmen served in Southeast Asia during the Vietnam War, but only seven were killed, reflecting the relative safety of their higher ranks. Nearly everyone in the class who remained in the military had been promoted at least to lieutenant colonel by 1970, the first year in which they were eligible for retirement with benefits. Mandatory retirement year (with few exceptions) was 1978 at the lieutenant colonel rank, 1980 at the colonel rank, and 1985 at all general ranks. Eight percent of the graduates of 1950 became generals (the first ones in 1971), making them by this measure one of the most successful of modern West Point classes. Most of these generals remained on active duty during the 1980s, filling the highest ranks of the American military.

Methods

DATA

New findings to be reported here depend on important improvements in method from that used by Mazur, Mazur and Keating (1984). Their original study was based on published sources of data. West Point's student yearbook, *The Howitzer* (1950), contained cadets' portraits and indirect measures of other variables including height and athletic prowess. The academy's annual *Register of Graduates and Former Cadets* (1964, 1980) listed records of promotion, but the 1980 edition of the *Register*, the latest available for that study, did not contain final ranks for the most successful graduates, those still on active duty in 1980.

These original data are merged with new data obtained for the present study from a questionnaire mailed in 1990 to 539 men from the class of 1950 for whom recent addresses were available. This was a fairly complete mailing to the surviving members of the class. After two follow-up mailings, we received 434 completed questionnaires, an unusually high response rate of 81%, no doubt because of the participation of West Point's Association of Graduates in the

project. Of 416 classmen who served 20 or more years in the military, 334 (80%) returned questionnaires, including virtually all the generals. The 16-page questionnaire asked about promotions, family background, education, career motivations, fertility, and family relationships.

VARIABLES

We are interested in predicting a classman's military rank attainment by the time of retirement, as reported in his questionnaire. Considering only the 337 respondents who served 20 years or more in the military, the distribution of final ranks is, in ascending order, major (1%), lieutenant colonel (26%), colonel (56%), brigadier (one-star) generals (7%), major (two-star) generals (5%), lieutenant (three-star) generals (3%), and "full" (four-star) generals (2%).

Elsewhere a considerable number of independent variables, including several measures of family background, were tested as predictors of military rank (Mazur, Mazur & Keating 1984; Mazur & Mueller in press; Mueller 1991).

Here we restrict our models to the following independent variables, which are of particular theoretical or empirical interest.

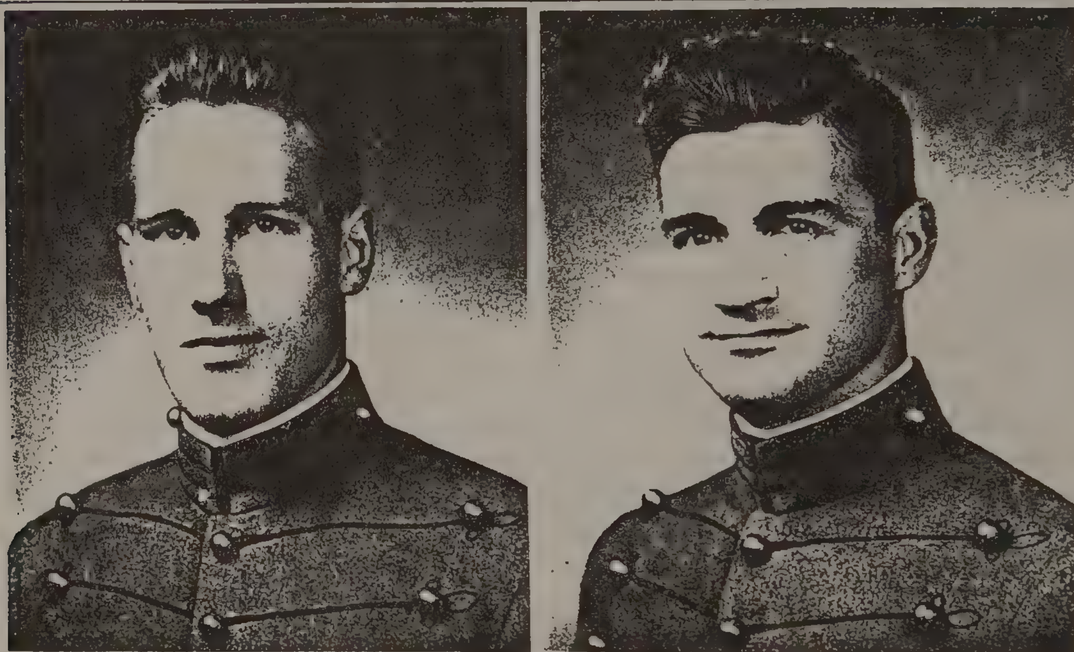
Face

Facial dominance was measured for all cadets who remained in the military for 20 years or more. Their yearbook graduation portraits (*Howitzer* 1950) were copied on slides for projection in front of 20 to 40 judges (usually undergraduate classes), who independently rated faces on a seven-point scale of dominance-submissiveness (1 = very submissive, 4 = neutral or undecided, 7 = very dominant). Judges were instructed that a dominant person tells other people what to do, is respected, influential, and often a leader; submissive or subordinate people are not influential or assertive and are usually directed by others. On 86% of the rated slides, at least half the judges' choices fell within two adjacent scale points, indicating more clustering than would be expected if the choices were uniformly distributed across the seven-point scale. Scores on the 14% of the slides that did not meet this cluster requirement were dropped as unreliable (the median of those dropped, however, did not differ from the median of the rest). The median score for each slide was taken as the value for *face*. These values ranged from 2 (moderately submissive) to 6 (moderately dominant) with a median value of 4 (neutral) and a mode of 5 (slightly dominant). Smiling, especially broad smiling, is known to lower facial dominance (Keating, Mazur & Segall 1981a, 1981b). In our sample, slight smile led to a 0.3 score point decline, broad smile (teeth visible) to a 0.7 score point decline. Dominance scores were adjusted accordingly (Mazur, Mazur & Keating 1984). Figure 1 displays cadet faces exemplifying a range of dominance scores.

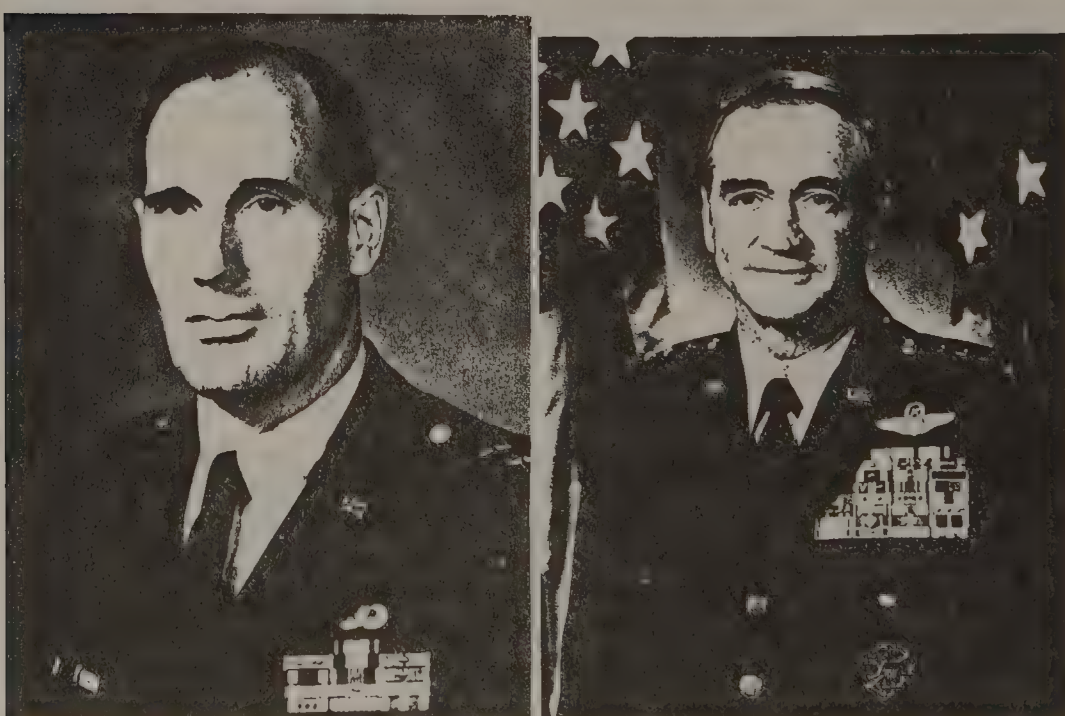
Athletic

By common standards, nearly all cadets are athletic and well proportioned physically, since these qualities influence admission to the academy. In order to mark the most athletic of these men, a dummy variable *athletic* was scored as 1 for the 16% of the classmen listed in *The Howitzer* as participants on a collegiate team in their senior (varsity) year; the remainder of the class was scored 0.

FIGURE 1: Cadets of Varying Facial Dominance Who Became High-Ranked Generals, as Cadets and at the Peak of Their Careers[a]

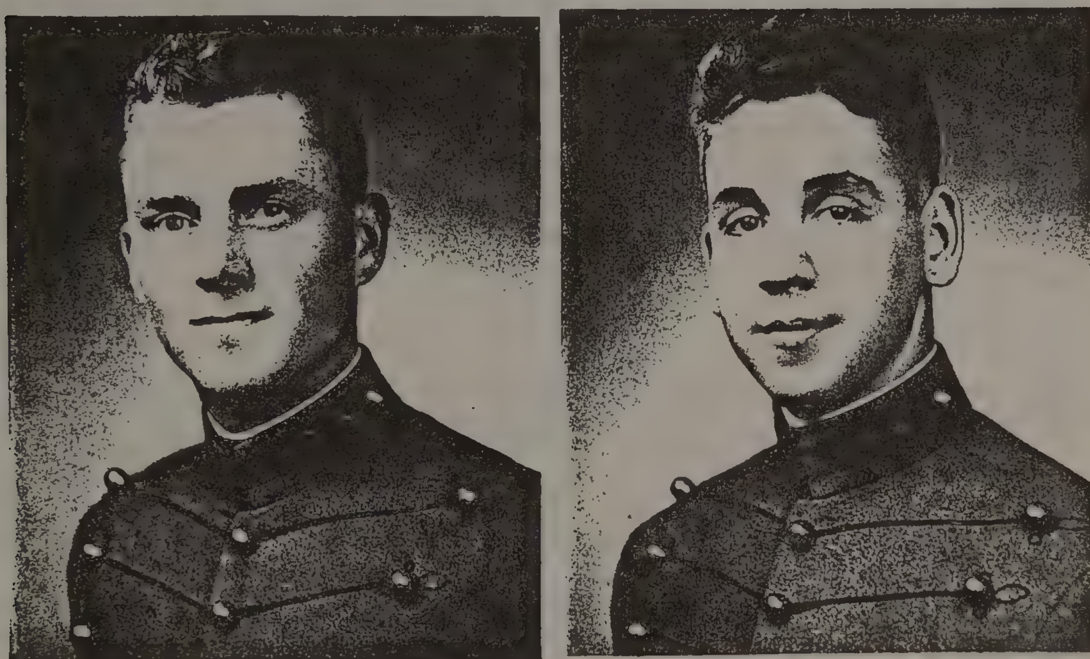


Wallace Hall Nutting | Charles Alvin Gabriel



[a] Faces are arranged from most to least dominant. The cadets with their ranks and final positions are, from left: General Wallace Nutting (*face* = 6), commander-in-chief, U.S. Readiness Command; General Charles Gabriel (*face* = 5), Chief of Staff, U.S. Air Force;

FIGURE 1: Cadets of Varying Facial Dominance Who Became High-Ranked Generals, as Cadets and at the Peak of Their Careers[a] (Continued)



John Adams Wickham Jr. Lincoln Faurer

[a] General John Wickham, Jr. (*face* = 4), Chief of Staff, U.S. Army; Lieutenant General Lincoln Faurer (*face* = 3), head of the National Security Agency. Portraits as cadets are from *The Howitzer 1950*, portraits as officers are from the U.S. Army Military History Institute and the Center for Air Force History.

Athletic is not simply a measure of physique, since it must include other personal and symbolic qualities associated with athletic prowess.

General Order of Merit

Until 1978, every graduate of West Point was given a number at graduation to indicate his "general order of merit" (GOM) within the class (*Register* 1980), the lower the better. Basically a performance measure, this aggregate evaluation combines academic grades, peer and instructor ratings of leadership and military aptitude, and physical education grades. At least until midcareer, GOM is known to be related to subsequent promotion (Butler 1976; Mazur, Mazur & Keating 1984).

Friends

Mazur, Mazur and Keating (1984) found that a crude measure of cadets' sociability predicted high military rank, so we include it here. There is a description approximately 50 words long of each graduating cadet in *The Howitzer*, typical of student yearbooks. The 41% of cadets whose descriptions made specific reference to their "friends" were scored 1 on the dummy variable *friends*; others were scored 0.

War College

Graduation from a war college is an essential step in becoming a general. Attendance at any one of five equivalent war colleges and date of graduation were coded from the *Register of Graduates* (1991). We do not treat war college attendance like the other independent variables, regarding it instead as a screening or interaction variable that specifies the pool from which generals will be drawn.

Parents' Education

Social class background, usually measured by parental education, is an essential feature of traditional models of status attainment. Here the father and mother are coded as having graduated from college or not. In addition, father is coded as having been a military officer or a West Point graduate himself.

Branch

Some of the army branches (such as infantry and armor) in combat can bring closer physical contact to an adversary than others (such as artillery, including coastal artillery, corps of engineers, signals, and ordnance) or the air force, and may, therefore, attract men who are more aggressive and, therefore, may look more dominant. In addition, career chances may vary among branches or services.

Analysis

We have nothing to add to Mazur, Mazur and Keating's (1984) results on early promotions, from West Point cadet officers up to the rank of major. Here we

analyze promotions to the middle and higher ranks: to lieutenant colonel and the various grades of general. We restrict our analysis to the classmen in our sample who remained in the service for 20 years or more, i.e., to those who made a life commitment to a military career. Nearly all (save four men) achieved the rank of lieutenant colonel or higher. Due to a lack of reliable dominance scores (see the discussion of *face* above) in 45 cases, we are left with a final sample of 289 career officers.

We recognize two aspects of promotion: (1) Does a man ever rise to a higher rank? (2) If so, how quickly does he do it? The two aspects are connected by the fact that in an organization in which promotion is based on merit, factors that make a promotion more or less likely may be the same as those that accelerate or slow down the speed of promotion.

Mazur, Mazur and Keating (1984) compounded these two aspects by comparing men who had attained different ranks by a given date (1964 and 1980). We use a different analytical strategy here. In general, there are two approaches to investigating the dynamics of these military careers. In one approach we treat the major transitions in a military career separately. For example, the promotion from lieutenant colonel to colonel was a virtual certainty for those who graduated from a war college. It was obtained by 53% of lieutenant colonels who did not. Being promoted to colonel for a nongraduate of war college is purely a reward — his career cannot proceed any further. Being admitted to war college, on the other hand, is not only a reward; it is an investment in that man. Less than half of the war college graduates, on the other hand, were promoted from colonel to brigadier general. Thus, each promotion poses different circumstances.

In the second approach, we investigate all of a man's promotions, counting from different base ranks onward; that is, we count total promotions above the level of colonel, brigadier general, etc. This can be justified on the grounds that the more capable a man is, the more often he will convince promotion boards, school selection boards, and command selection boards of his worth. Thus, subsequent promotions can be viewed as cumulative steps in one's career process and may be treated as an interval measure of status attainment from a given level.

Both approaches pose methodological problems.

In the first approach, if we investigate each promotion independently, we could model it as a transition process with two risks: promotion and retirement. The problem here is that the two risks are not independent, the censoring process — premature retirement — eventually depending on the distribution of the promotion risk. Officers are well aware of their chances of obtaining another promotion.

A man of the class writes about the promotion to colonel: "You are considered each time the promotion board meets for the next rank. And, since the promotion board normally meets once a year, you are considered each year. . . . If you are passed over to lieutenant colonel or to colonel . . . , in effect, you are dead. Forget the next two years. . . . For instance, the most common passover to colonel is caused by a poor efficiency report (maybe just "excellent" instead of a "max") as a battalion commander. Everything else was okay, or you wouldn't have made lieutenant colonel. During the next 2 years, nothing you

can do will change that lousy command OER (Officer Evaluation Record) in your file. Your best bet: retire."

Another classman writes about the promotion to brigadier general: "Not only is the promotion board to brigadier general given strict criteria, but all colonels in the 'zone' are aware of the criteria which [are] always published. . . . I think I was competitive for brigadier general for about 1 year, during and just after my tour as XXX. . . . After the next board met, and I wasn't selected, I knew that my chances were zero."

We may conclude from these statements that most men who retired before they reached the mandatory retirement age of their rank (28 years of service for lieutenant colonels, 30 years for colonels, 35 years for all grades of general) did so, among other reasons, because they — more or less correctly — assessed their chances of promotion to be low. We may assume that they would have stayed longer had their chance of promotion been better. On the other hand, many men with little or no chances of promotion may have stayed until the end because of a lack of attractive alternatives. The interdependence of the two risks is, however, one-sided: the probability of premature retirement may have been influenced by perceived promotion chances, but not vice versa. Actually, none of the factors considered here as affecting chances of promotion on any level was a predictor for premature retirement.

The alternative, modeling the process as a one-risk process means effectively treating all men who were not promoted as if they had stayed until mandatory retirement age. Since every man in a given rank is being considered for promotion throughout several years, application of a mover-stayer model — assuming unobserved heterogeneity with respect to promotion chances — would be highly questionable, especially if we assume that, in the end, all movers have moved and all stayers have stayed. On the other hand, neither is the assumption plausible that all men are movers, some of whom, however, were not allowed to move but might have moved had the demand for officers on the next rank been greater at the time. Even in wartime there will always be a substantial proportion of men in any rank who will not be promoted further. Taking all things together, modeling career transitions as a process with two independent risks might be a lesser distortion of reality after all. In fact, the estimated factor effects were not different for the one-risk and the two-risk models.

A third possibility would be simply to compare all those who were promoted with those who were not by a multivariate logistic regression model. By doing so we would circumvent the problems caused by the intricate censoring process but would give away information about the time dimension of the process: the speed of the career. As for many other life events, the same factors that increase the chances of promotion may accelerate its occurrence. These factors also may increase the chances of subsequent promotions.

For the second approach, the choice of an appropriate model is not obvious either. Since, here, we are not interested any more in transitions from or to a single state = rank, but rather in final career outcome, total number of subsequent promotions, as seen from some rank level already achieved, would be the dependent variable. All possible promotions take place within a fairly similar time span — all promotions save one to brigadier general and 80% of all

promotions to major general took place before 1978, when lieutenant colonels had to retire; all promotions save one to major general and 80% of all promotions to lieutenant general took place before 1980, when colonels had to retire; general officers of all ranks could stay until 1985. Thus, total number of subsequent promotions is a very good indicator not only of final outcome but also of speed of career.

The appropriate model for analyzing these count data, however, is not a point process (Cox & Isham 1980), because after any promotion there is an official minimum waiting time of two years (there may be exceptions for the highest ranks), a substantial proportion of total average waiting time during which no promotion can occur, but premature retirement can; and because the number of events that one person can experience is limited: no one can obtain promotion beyond full general.

We chose the following strategies:

We investigated single transitions with a two-risk (promotion and retirement) model, but we look only into the factors affecting the risk of promotion. Reasons when to retire — see quotations above — may entail, in addition to the perceived promotion chances, a variety of factors from family situation to labor market conditions that lie outside the focus of this study. In fact, none of the factors considered in this study predicted time of retirement. Since the underlying time process (annual promotion board meetings) is discrete, there is no theoretical reason to prefer one parametric model over the other. In particular, there is no reason to assume that one parametric model is the appropriate one for all transitions considered here. Therefore a proportional hazard rate model as described by Cox (1972; Cox & Oakes 1984) is the model of choice. We will consider graduation from war college, promotion from major to lieutenant colonel (which all save four men in the sample obtained), from lieutenant colonel to colonel, from colonel to brigadier general and from brigadier general to major general. The last promotion is tantamount to accession to the military elite nucleus. The crucial condition of proportionality of the influence of covariates is fairly well fulfilled (see the discussion of results).

In addition, we compared with a logistic regression those who made the transition with those who did not.

For analyzing the total number of promotions as seen from various ranks onward we simply chose multivariate OLS regression with the log of total number of promotions as the dependent variable. For the interesting ranks (colonel with war college, brigadier general, major general — all lieutenant colonels with war college became colonels, all without could not become more), the total number of promotions possible range from 0 to 4, 0 to 3, and 0 to 2, respectively. For that many levels in the dependent variable, OLS methods can give sufficiently robust results (Bentler & Chou 1987).

Results

Measures of parental education including father's military background are unrelated to any career variable (same findings in Butler 1971b and Janowitz 1971) and were therefore dropped from subsequent analysis.

TABLE 1: Correlations (Gamma) between Dichotomized Independent Variables

	Athletic	General Order of Merit	Friends	War College
Face	.11	.12	-.04	.01
Athletic		.12	-.17	.22
General order of merit			.05	-.42**
Friends				.06

(N = 289 to 334)

* $p < .05$ ** $p < .001$ (chi-squared test)

Career chances after graduation from war college were better for classmen in the air force than in the army (17 out of 27 became generals, as compared with 39 out of 113), perhaps because, in the post-Vietnam era, the air force was not cut back in size as much as the army. No variable considered here nor any other data obtained in the survey predicted a man's entering the air force. Also, all effects on career dynamics reported below could be observed for air force and army alike. Classmen who entered "combat" branches within the army (infantry, armor) had the same facial dominance scores as men opting for one of the other army branches. No such differences could be observed among those still on active duty 20 years after West Point, nor among war college graduates, nor among generals. Likewise, career chances were equal for all army branches — a closely watched issue for officers and personnel planners anyway. Consequently, membership in a particular army branch or the air force was dropped from further analysis.

Table 1 contains a correlation matrix of the remaining independent variables using gamma, which is insensitive to marginal distributions. Variables are dichotomized as indicated above or, in the case of *GOM* and *face*, at their medians. The only significant gamma is between *GOM* and *war college*, indicating that men with lower (i.e., better) GOM rankings were more likely to graduate from war college. Otherwise there is no multicollinearity to complicate the interpretation of regression coefficients.

DISCRETE PROMOTIONS

Graduation from war college is a prerequisite for being promoted to general and virtually guarantees promotion to colonel. Admission to war college effectively separates officers into two career "channels" (a concept we have developed elsewhere: Mazur & Mueller in press): the one eventually leading to general's stars, the other not.

TABLE 2: Coefficients of Dichotomous Covariates for Transition to Ranks from Rank Below and to War College Graduation according to the Proportional Hazard Rate Model

	Transition to								
	Major	Lt.	War Coll. Grad.	Col.-	Col.+	Brig. Genl.+	Major Genl.+	Lt. Genl.+	Genl.+
Indicator variables									
Face	.02	.03	.08	-.21	.12	.25	.59	.30	1.44
Athletic	.03	-.01	.19	.32*	.22*	.18	.67	1.44	-1.73
General order of merit	-.08	-.11*	-.36**	-.18*	-.04	-.27	.10	.45	-.68
Friends	.03	.10	.08	.17	.38	.44	.60	.59	2.82

+ War college graduates only
- Nongraduates of war college only

* $p < .05$ ** $p < .001$

Following our hypothesis that facial dominance might — deliberately or not — be regarded as an important factor for any high leadership position but irrelevant for "middle management" positions, we expect that *face* may influence a man's chances to be admitted to war college, and for the sequence of promotions from lieutenant colonel to four-star general in the channel opened by that admission, but may be irrelevant for the promotion from lieutenant colonel to colonel for those men who did not go to war college.

Accordingly, proportional hazard rate models were estimated for admission to war college and for promotion from major to lieutenant colonel for all men in the sample, and for all five promotions from lieutenant colonel to four-star general in the channel for war college graduates. In addition, a proportional hazard rate model was estimated for the promotion from lieutenant colonel to colonel for those men who did not go to war college. Base line estimates were made according to Breslow (1974); given the lack of a generally accepted goodness-of-fit test statistic for these models (Andersen et al. 1992:208-15; Fleming & Harrington 1991:92-94; Lin, Wei & Ying 1993; Lin & Wei 1991) we confined ourselves to graphical testing (see below); the program was TDA written by Goetz Rohwer of the University of Bremen.

The effect of *face* in the channel for war college graduates is not significant for any single transition, but the effect is always in the predicted positive direction (sign test: $p < .05$). *GOM,* while being a powerful predictor of graduation from war college, shows an inconsistent effect otherwise, as does *athletic.*

Having had *friends* as a cadet displays the same consistently positive effect on transition rates as *face* in this channel. *face,* on the other hand, has no effect on promotion from lieutenant colonel to colonel in the other channel (see Table 2).

TABLE 3: Survival Functions, Beginning in Year of First Observed Transition[a]

	To Lieutenant Colonel		War College		To Colonel-		To Colonel+	
Year	Sub.	Dom.	Sub.	Dom.	Sub.	Dom.	Sub.	Dom.
1	.923	.942	1.000	.988	.991	.986	.986	.973
2	.794	.785	1.000	.982	.991	.986	.709	.695
3	.458	.434	1.000	.977	.991	.986	.525	.415
4	.434	.427	.994	.970	.954	.986	.167	.113
5	.111	.414	.965	.888	.935	.905	.080	.038
6	.099	.101	.829	.809	.703	.650	.050	0
7	.034	.094	.728	.733	.425	.454	.050	
8	.016	.017	.665	.675	.391	.431	0	
9	.020	.011	.622	.602	.391	.431		
10	*	0	.574	.584	.391	.431		
11			.565	.556	.391	.431		
12			*	.534	.353	.358		
13				*	*	*		

[a] Graduation from war college: all men serving for 20 years and more; promotions: all men in rank below, calculated for actual risk population — excluding prematurely retired men.
Submissive (Sub.) = *Face* score at or below median;
Dominant (Dom.) = *face* score above median.

\+ War college graduates only
\- Nongraduates of war college only
* No further transitions

The estimated survival functions for all five transitions in the career channel for war college graduates, comparing men with *face* scores above the median with those below it, show that in every rank, at nearly every point in the respective interval, men with high *face* scores were subject to higher transition rates than men with low scores (the difference between proportions surviving is monotonically increasing, as required by the proportional hazard rate model). Given the strong positive correlation between early promotion to a rank and the chances of another promotion, the differences between single survival functions even understate the real advantage high *face* scores secured to men in this career channel. For promotion from major to lieutenant colonel, for admission to war college, and for promotion from lieutenant colonel to colonel for men who did not attend war college, this does not apply (see Table 3).

In addition, we used logistic regression analysis of discrete promotions of single transitions, comparing all those who experienced the transition with those who did not.

Since all save four men in the sample obtained promotion to lieutenant colonel, and since graduation from war college virtually assured promotion to

TABLE 3: Survival Function for Transition to Higher Rank and for Graduation from War College (Continued)

	To Brigadier General+[b]		To Major General+		To Lieutenant General+		To General+	
Year	Sub.	Dom.	Sub.	Dom.	Sub.	Dom.	Sub.	Dom.
1	.985	.968	.963	1.000	.938	1.000	.857	1.000
2	.924	.903	.704	.714	.938	.889	*	.700
3	.833	.774	.587	.500	.813	.778		.583
4	.724	.661	.547	.321	.750	.500		.467
5	.712	.597	.508	.286	.600	.444		*
6	.636	.581	.469	*	.600	*		
7	.621	*	*		.480			
8	*				*			
9								
10								
11								
12								
13								

[b] Three distinguished professors at the academy or other military colleges were promoted brigadier general upon retirement in the late 1980s. These promotions were not considered here.

colonel and is an essential step toward higher rank, at midcareer we focused on admission to war college from any rank (usually lieutenant colonel). GOM is the only independent variable significantly related to war college admission (logistic partial correlation coefficient $R = -.21, p = .001$). Also, *athletic* ($R = .13, p = .003$) and *GOM* ($R = -.07$, $p = .04$) were moderately strong predictors of promotion to colonel for men who did not attend war college. Neither *face* nor *friends* played any role for these two transitions.

Since only war college graduates become generals, our regression for promotion from colonel includes only men who graduated from war college ($N = 140$). Among these, only *face* is a significant predictor of promotion from colonel to brigadier general ($R = .12$, $p = .05$). Among brigadier generals, *face* is the only predictor of promotion to major general that approaches significance ($R = .12, p = .10, N = 51$). The number of men at even higher ranks is too small to expect significant effects, and there are none for promotion from major general to lieutenant general ($N = 31$). However, for promotion from lieutenant general to full general ($N = 16$), both *face* ($R = .09$, $p = .13$) and *friends* ($R = .19$, $p = .09$) approach significance.

NUMBER OF PROMOTIONS

Figure 2 shows the average number of promotions (from a base rank) as a function of facial dominance. There are three curves, corresponding to three base ranks: colonel, brigadier general, and major general. For example, the solid line shows the average number of subsequent promotions for colonels at each value of *face*. Clearly, the more dominant a colonel's face, the more future promotions he enjoyed. The other curves give a similar picture. (No men with *face* values below 3 were promoted to general, so the graphs showing promotions from brigadier general and major general begin at *face* = 3.) The graphs show men with *face* scores of 6 receiving .3 to .8 more promotions than men with *face* scores of 3.

The number of promotions from the ranks of colonel, brigadier general, and major general was regressed on the independent variables. For example, for all men who reached the base rank of colonel, we used OLS regression to predict their number of future promotions. Standardized coefficients (betas) and their significance levels are shown for each base rank in Table 4. Since men who do not graduate from war college by the time they are colonels have no chance of further promotion, we include only war college graduates in the regression from the base rate of colonel. Of 140 colonels who had graduated from war college, only 56 men reached brigadier general and 31 attained major general rank, and so significance tests are not equally powerful across these regressions.

Regression results in Table 4 vary considerably across base ranks. The same four independent variables explain more variance (R^2) in promotions from brigadier general and major general than in promotions from colonel.

Face is a strong and significant predictor of the number of promotions from colonel and becomes an even stronger predictor of future promotions for brigadier generals. *Face* remains a sizable predictor of promotions from major general although there are only 31 subjects in this final regression.

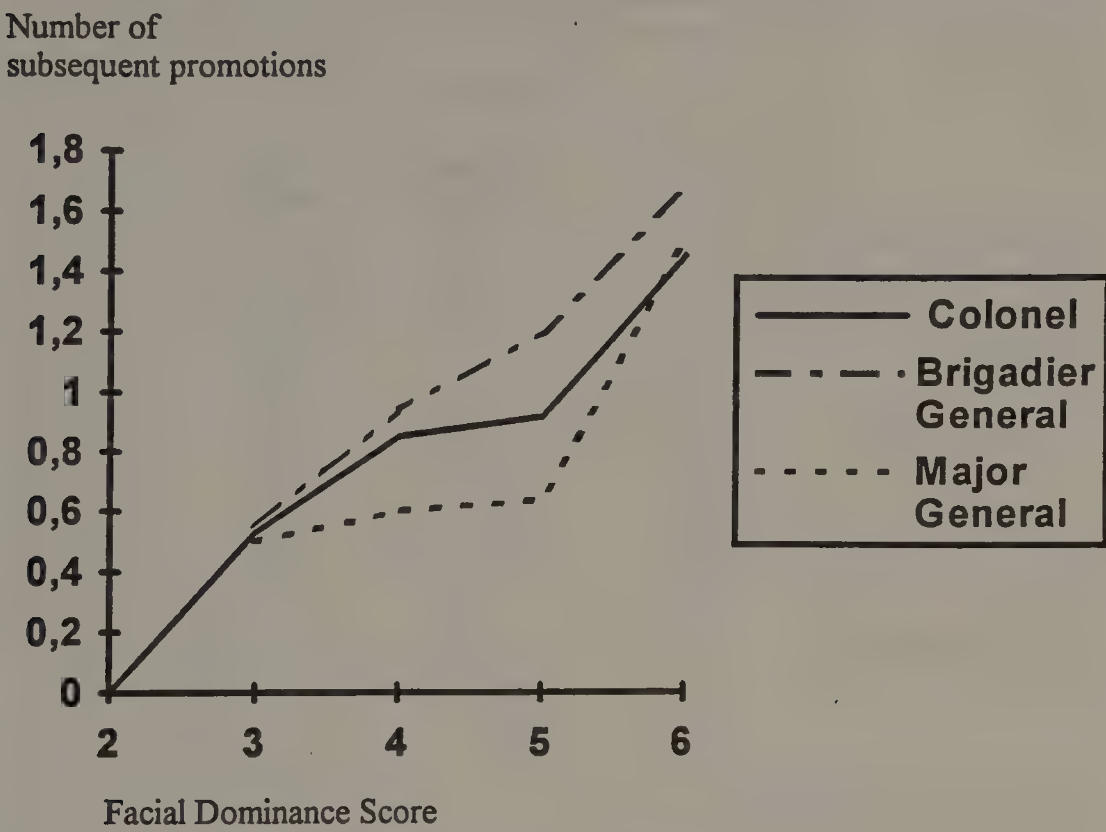
Varsity athletics (*athletic*) approaches significance in predicting promotions from brigadier general ($\beta = .24$), but given the small number of brigadier generals, its effect does not reach significance ($p = .07$). It is neither a strong nor a significant predictor of promotions from colonel or major general.

GOM has little predictive power when the promotion pool is limited to war college graduates. A reason could be that men with better GOM are more likely to enter war college (Table 1), thus restricting the range of the variable. This effect, however, if any, is weak: lower quartile and median and upper quartile of GOM for all classmen as well as for all lieutenant colonels was 48, 64, 83; for all war college graduates, 44, 58, 73; for all generals, 44, 57, 71.

Friends is a modest predictor of future promotions from all base ranks, significant or very nearly significant for promotions from colonel and brigadier general. Considering the extreme crudity of this measure of sociability, it is noteworthy that it shows any effect at all.

Together, the independent variables explain more variance in promotions once the pool is limited to war college graduates. While *GOM* loses its predictive power here, *face* and to a lesser degree *friends* become more important. *Face* is a sufficiently strong predictor for the career potential of colonels and

FIGURE 2: Number of Promotions from Each Rank, by *Face*[a]



[a] 20+ years of service, war college graduates only. Values for *face* = 2, for *face* = 6 from brigadier general, and for *face* = 3 and 6 from major general, are based on very small samples of size 4-9. Remaining values for brigadier general and major general, and for *face* = 6 from colonel, are based on samples of 10-19. Remaining values for colonel are based on samples of 103.

brigadier generals that it reaches significance despite the small sample. Figure 2 leaves little doubt that *face* is important in promotion from major general, too.

Event history analysis and logistic regression analysis of discrete promotions are less powerful and show fewer significant effects than OLS analysis of total promotions, but the single event effects that do appear are consistent with OLS results. Both point to *GOM* as an important predictor of admission to war college but not of later advancement within the pool of war college graduates. Both show *face* to be unimportant at middle rank, but after graduation from war college, both show *face* to be consistently the most important predictor of high rank.

TABLE 4: Ordinary Least Square Regressions for Number of Promotions (Log Transforms), from Base Rank, for War College Graduates[a]

	Base Rank		
	Colonel	Brigadier General	Major General
Indicator variables			
Face	.16	.26	.23
	(.06)	(.05)	(.22)
Athletic	.11	.24	.18
	(.18)	(.07)	(.34)
General order of merit	-.04	-.04	.14
	(.59)	(.77)	(.48)
Friends	.17	.26	.21
	(.04)	(.05)	(.24)
R^2	.09	.19	.19
Model significance	(.02)	(.03)	(.24)
N	118	51	31

[a] p values are in parentheses.

Discussion

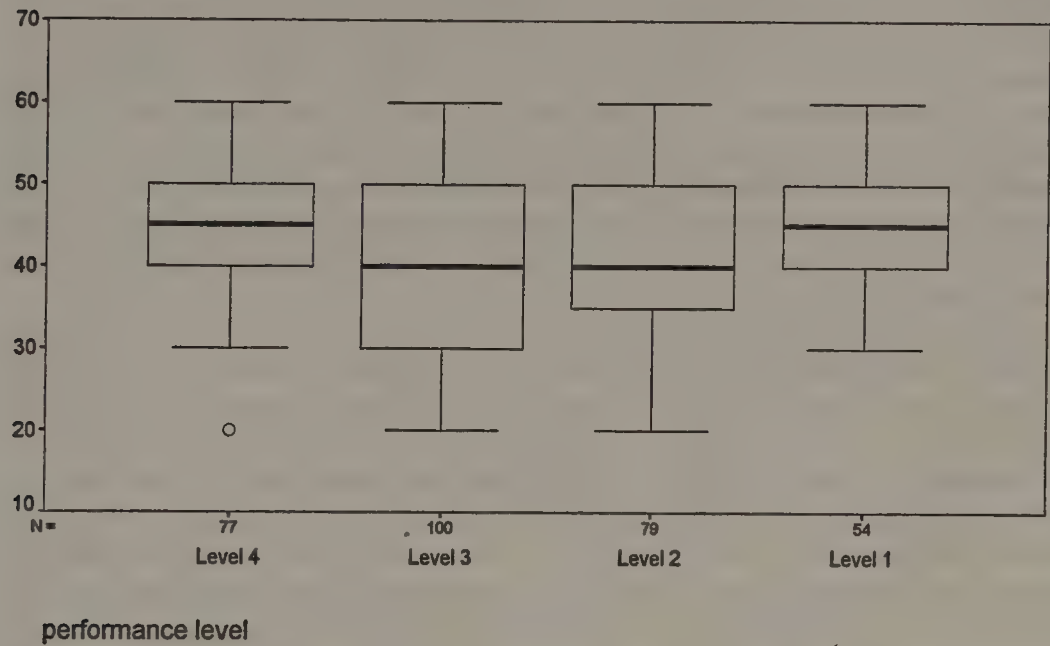
Improved methods show that cadets' facial dominance, measured from their West Point graduation portraits, predicts military rank 20 years or more later. The finding is theoretically important in part because it allows us to evaluate hypotheses raised by Mazur, Mazur and Keating (1984) at the conclusion of their research.

First, we can reject any suggestion that dominant appearance in a graduation portrait is the effect, rather than the cause, of attaining the status of cadet officer while at West Point. The fact that these portraits correlate with rank 20+ years later, as well as at West Point, while the status as cadet officer does not correlate with rank in late career (Mazur, Mazur & Keating 1984; see also Janowitz 1971), makes the causal direction unambiguous.

Second, we can also reject the hypothesis that in middle age, physical features lose their relevance for hierarchical placement.

Third, it was also hypothesized that *face* failed to predict future rank because middle-aged men no longer look like their graduation portraits. It is certainly true that looks change with age. But since *face* does predict the highest level promotions, its failure to predict earlier promotions must be due to some other factor than the timeliness of the portrait. This conclusion is buttressed by a supplementary study of 30 classmen who provided us with portraits taken in

FIGURE 3: Facial Dominance by Performance Level



Performance levels:

Level 4: no war college; final rank = lieutenant colonel
Level 3: no war college; final rank = colonel
Level 2: war college graduates; final rank = colonel
Level 1: war college graduates; final rank = general

middle age (average year = 1972) that could be compared to their cadet portraits (Mazur & Mueller 1993). Middle-age portraits were scored for dominance by the same method used on the cadet portraits. Although poses, dress, and formats of the middle-age portraits are considerably more variable than those of cadet portraits, the correlation between dominance ratings of portraits taken 20 years apart is gamma = .44 (p = .04). Men who looked dominant as cadets tended to look dominant in middle age as well. This finding accords with the stability over a lifetime reported for other facial impressions like babyfacedness and attractiveness (Zebrowitz & Montepare 1992; Zebrowitz, Olson & Hoffman 1993).

Fourth, we can also discard the hypothesis that the absence of an effect of facial dominance on early and middle career is due to a sampling bias: that the uninterrupted effect of facial dominance was masked by the — on the average — greater facial dominance of West Point graduates than of their far more numerous competitors who obtained their commission elsewhere.

Apparently the percentage of West Point graduates among war college graduates was not much lower than the percentage of West Point graduates among generals: that is, in the 40-50% range (Atkinson 1989, quoting official sources; Moore & Trout 1978). On the other hand, for the men in our sample, the most restrictive filter in the long sequence of career steps was admission to war college: of those eligible, only 39% passed it (only the last filter of all, the one to full general, with a 35% chance for those eligible, was even harder to pass). If facial dominance were important also in midcareer, then, even if West Pointers look more dominant than other candidates, its effect should have become visible in our sample, at least at the transition to war college. Since it was not, we conclude that the effect existed only at the academy and at the top of the hierarchy.

This leads us to favor the remaining hypothesis, that facial appearance is a leadership quality that is more important for some promotions than for others because it signals qualifications that are more important for winning in primary group contests. Consequently, it should be more important when those recommending promotion know the candidate personally. Such is the case both at West Point, where the decision is based on evaluations by cadets and tactical officers, and in promotion to the highest ranks, attained by a few men toward the end of their careers, when the small pool of candidates is again personally known to those in judgment. At midcareer, their physical appearance is not an important factor compared to their records, even though the portraits accompany promotion files.

Academic attainment, as measured by GOM, is the only one of our four independent variables that is conventionally regarded as an important determinant of promotion in merit-based bureaucracies, which the modern military certainly is. It is puzzling that at the highest ranks, this indicator of technocratic competence loses its prior relationship to promotion. Since the range of GOM is only slightly restricted by the sequence of filters in this channel, we may conclude that after graduation from war college, for high positions in the hierarchy, leadership qualifications not measured by GOM become more important. At the same time, factors that seemingly ought to be irrelevant in a meritocracy — facial dominance, athletics, and friendliness — increase their relationship to promotion.

The best-measured of our four independent variables, GOM, is based on four years of college achievement, yet it is unrelated to promotion to the highest ranks. In contrast, the other variables are operationalized in ways that seem barely adequate: *face* is measured from a yearbook portrait taken 20 or more years before the highest rank is attained; *friends* is coded from a 50-word profile of each cadet in the yearbook; and *athletic* measures some ambiguous combination of physique, athletic prowess, team spirit, and in the case of lacrosse or football, school glory. It is remarkable that they relate to anything, yet these are the variables most predictive of final promotions. It is a compelling hypothesis that improved measures of these variables would produce far stronger relationships than are reported here. If true, then the effect of these variables on promotion is substantial.

What is the social mechanism behind the positive effect of facial appearance on status attainment at the academy and at high levels of the military hierarchy, but not at the low and middle levels?

Virtually all studied species of nonhuman primates use facial gestures to signal social status (dominance and submissiveness) (Smuts et al. 1987). These gestures help to regulate relationships among conspecifics by settling conflicts without violence or with limited violence. Homologous human facial gestures have been shown to exist and to be reliably and uniformly interpreted in many cultures (Keating, Mazur & Segall 1981a, 1981b). Characteristically, conflicts of all sorts are settled by an escalation of warnings, threats and the application of force with increasing intensity: the outcome is decided once one of the contestants decides not to move up the escalation ladder any further. Conflict resolution thus crucially depends on the perception and credibility of escalation threats. Cues for the credibility of escalation threats may be called signals.

Facial dominance as analyzed in this study seems to signal the potential for as well as the intention of escalation (for a detailed discussion of facial dominance as a biological signal see Mueller & Mazur 1995; see also Mazur 1985): strong jaws may indicate a heavily built skeleton, and therefore superior physical strength. A broad prominent forehead with strong eyebrow arches may indicate enhanced ability to absorb hits. The conventional signals that can be found in dominant looking faces seem to include several gestures observed in many nonhuman primates and humans as well when preparing for a fight (Harper 1991): thin lips, withdrawn corners of the mouth, lowered eyebrows with partially closed eyes (in order to protect them against injuries), withdrawn ears, making them appear smaller. Conversely, people with facial features typical of infants (large eyes, high thin eyebrows, round face, small nose bridges — "babyfacedness") are perceived as warmer, weaker, submissive (Berry 1990). Babyfaced people describe themselves as less aggressive (Berry 1991). At present, however, our knowledge is very limited about the functional contribution of such anatomical and gestural features to the overall perception of facial dominance. Also, for our sample we have only the scores of perceived facial dominance from the ratings.

Our data suggest that facial dominance did not function as a relevant qualification for junior or middle rank positions nor for admission to war college (as Collins & Zebrowitz 1995 in a small sample of men who had served in the military for some years — mainly from age 20 to age 25 — found no correlation between babyfacedness and number of promotions), but did so function for promotion to cadet ranks at the academy and for promotion to the top positions in the military hierarchy. There must be some relation between dominant looks and the abilities it signals. If the selection of persons for an important position is made dependent on many sensible indicators for competence, then adding a criterion that is completely uncorrelated with competence will inevitably decrease the average quality of the selection. Therefore, if selection to the top ranks is made also on grounds of perceived facial dominance, facial dominance must be a useful feature of a cadet or a general.

These findings still leave room for various interpretations of the social meaning of this set of basically biological signals, which evolved in an environment of long ago.

Typically, conflicts within modern organizations are not decided by threat or application of physical force; such behavior is severely sanctioned. Therefore, along the line of the "novel environment" approach (Barkow, Cosmides & Tooby 1992), it could be argued that while there is a universal, even biological, basis for facial dominance signals, in the social reality of modern meritocratic organizations they carry some other meaning.

Perhaps promotions into the military elite, the top levels of the hierarchy, go preferentially to those who present themselves as tough warriors, even if the management of modern warfare requires quite different abilities. Embodying the ideals of the hierarchy in such a direct way may also help to prevail in primary group contests within the organization. This would be an application of Moore and Trout's (1978) "visibility" theory in a symbolic interactionist tradition: Using a biological communication channel in order to send a culturally mediated message, men present themselves in conformity with the values of the established hierarchy that they want to join. Several empirical findings in our sample, however, make it difficult to defend this interpretation.

Such an interpretation first must assume that, in this particular professional group, facial appearance is prized and possibly rewarded. This much definitely is the case — to men in our sample the findings of this study neither came as a surprise nor as an offense (see also the opening discussion of this article).

On the other hand, a smile, in particular a broad smile, decreases the facial dominance score awarded by raters and, therefore, probably decreases perceived facial dominance, too. Mazur, Mazur and Keating (1984) found that broad smile (with teeth visible) decreased facial dominance score by 0.7 points (see the discussion of variables). The portraits in the graduation yearbook were well prepared and carefully taken, and furthermore, we may assume that, as in other American colleges, graduates had a chance to select their portraits for publication from several photos taken. We should then expect that ambitious men, those who had already attained a cadet officer rank, as well as those who wanted to rise high, were more inclined to present themselves on the portraits to be as dominant as possible and therefore should have avoided posing or selecting portraits with a broad smile altogether. There were however, no differences in the distribution of broad smiles over cadet ranks. Cadet officers displayed broad smile not less frequently than cadet sergeants (see Table 5). Likewise, there was no difference between future officer ranks: future lieutenant colonels, colonels, and generals all had displayed broad smiles in roughly the same numbers (see Table 6).

Also, if facial dominance of cadets as measured here indicates an individual's conformity with the values of the hierarchy, then, in addition, we should expect a correlation with *GOM* in the predicted direction, which could not be observed (Table 1).

In addition, presentation of oneself as being in conformity with the values of the hierarchy is a feature that will be valued for all promotions. If facial dominance is valued as indicating a candidate's devotion to the values of the military hierarchy, then, by all that we know about socialization in professions and organizations, we should expect that facial dominance should also influence one's chances for promotion during early and middle career. Candidates have to provide portraits, which are included in the files open to inspection by

TABLE 5: Smiling in Portraits of Men by Cadet Ranks in Senior Year[a]

Cadet rank	No or Slight Smile	Broad Smile
Captain	30	10 (25%)
Lieutenant	84	28 (25%)
Sergeant	412	105 (20%)
All	526	142 (21%)
(N = 669)		

[a] Data from *Howitzer*

promotion boards for all ranks. Again, the most ambitious candidates should have provided more dominant portraits in order to advertise their adherence to the values of the hierarchy, and it should have worked. If, besides smile, official portraits leave a man only little room for varying his facial dominance, then, given the stability of facial dominance scores through the life course, we should expect a positive effect of facial appearance as rated from the graduation portraits, on promotions throughout the whole career, which could not be observed. But if an officer, anticipating the expectations of promotion boards, can substantially vary his facial appearance on official portraits, why should portraits in early and mid career, but not in late career, differ in this respect from the portraits taken at the academy?

Therefore we are left with the alternative interpretation: that facial dominance also in the context of the modern military measures what the raters were asked to assess: the capability to actually dominate other people, even if this ability in the social context must be applied only in a subtle and purely symbolic way.

There are many instances of contests, both intraspecific and interspecific ones, where it pays to fake. All other things being equal, an intruder often will defer to the owner of a territory who has already invested in it and for whom it is more valuable; a well-fed individual will defer to a hungry one. This should be beneficial to both sides, if such differences in intentions and interests are honestly signaled in order to avoid costly fighting. But there is the option to fake: pretend ownership or hunger. But then, why should an opponent believe such a signal?

When it comes to signaling relative contest abilities within a rank order, first Zahavi (1977), and in two rigorous papers Grafen (1990) and Enquist (1985) — see also Hasson (1994) — showed with game theoretical models that honesty can be an evolutionarily stable strategy: a strategy that in any real context will displace alternative strategies and cannot be displaced by them. There is only one condition for this finding to hold: the signaling itself must be costly and it

TABLE 6: Smiling in Portraits of Men by Future Final Rank (Lieutenant Colonel Rank and Above)[a]

Final rank	No or Slight Smile	Broad Smile
General	43	14 (25%)
Colonel	173	55 (24%)
Lieutenant colonel	112	27 (19%)
All	328	96 (23%)
(N = 424)		

[a] Data from *Howitzer* and *The Register*

must be more costly to signalers who actually are less able to fight. Several experimental observations are consistent with this prediction: for example, subdominant male house sparrows with experimentally enlarged badges (a signal of dominance) are attacked by birds with naturally large badges more frequently than other subdominant individuals (Møller 1987).

Among humans, signaling a contest ability that does not exist can also be punished very effectively. Many communities where rank is formally signaled by uniforms, titles, badges, etc., have strict punishment for fakers; there are equally harsh sanctions against individuals who informally signal a higher rank than they actually have.

In our sample, there are clear hints that signaling facial dominance without fulfilling the according expectations indeed was costly. Among men who did not graduate from war college and who had a GOM score in the poorer half (above 72), of all factors considered here facial dominance together with GOM was the most potent predictor for promotion to colonel, but — unlike GOM — a negative one ($R = -.13$, $p = .04$, $N = 76$); neither *athletic* nor *friends* had an effect in this particular subgroup. The higher the facial dominance score, the lower the chances for promotion to the highest rank obtainable for that subgroup (low grades at the academy; no war college), which in terms of professional performance was the weakest quarter of the whole sample. Some of these men, had they looked less dominant, might have risen higher than they actually did. Survival function differences in Table 3, while not approaching significance, accordingly speak more for a negative rather than a positive impact of facial dominance on chances of becoming colonel for nongraduates of war college in general.

Also, if we look only at final rank, while facial dominance scores rose from medium performers (colonel, no war college) over good performers (colonels who had graduated from war college) to top performers (generals), those men who did not graduate from a war college and were never promoted beyond lieutenant colonel looked almost as dominant as generals (see Figure 3).

It is interesting that, in this low-performing subgroup, year of promotion to lieutenant colonel highly correlated with facial dominance: the higher the facial dominance score, the later this last promotion had occurred (Spearman $r = .37$, $p = .001$). This was the only instance in this data set in which a strong significant relation could be observed between facial appearance score and a parameter of career timing.

We may conclude from these observations that if a man visibly failed to meet the expectations of his superiors and peers — that a West Point graduate excels in professional performance — and instead displayed only modest capabilities, then looking like a dominant military leader was a handicap rather than an asset. Parallels can be seen in that babyfaced individuals, who generally are perceived as warm, dependent and submissive, receive harsher punishment for wrongdoing (Berry & Zebrowitz-McArthur 1988; Zebrowitz, Kendall-Tackett & Fafell 1991). Furthermore, our finding is another strong piece of evidence against the "visibility" hypothesis: presenting oneself as conforming with the values of the hierarchy should never be a handicap.

Here we have the relative costs of overstating one's contest abilities that the theory requires for honest signaling as an evolutionarily stable strategy. Thus, facial dominance, as measured by the raters and perceived by the promotion boards, rather than being a symbol for something else, is perceived also in the context of a military career as honestly signaling individual contest capabilities; higher facial dominance scores indicating a greater probability to win in a contest against men with lower scores. In midcareer, at the ranks of lieutenant colonel and colonel, expertise, cognitive competence, and social intelligence are the required leadership qualities. On the top levels of the military hierarchy more archaic features of leadership matter, too. Selecting cadet officers at the academy also by the criterion *face* can be seen as a purposeful preparation for that dimension of leadership.

Is the military, with its warrior traditions, unique in its attention to dominant physical features? Perhaps, but there is reason to think that military promotion *understates* physical effects. Because of selective recruitment and socialization, West Point cadets look physically more impressive in face, physique, and posture than do students at other universities, and the same may be said of military officers compared to civilian professionals. Thus, there is less variation in appearance among military than among nonmilitary populations, and therefore less opportunity for physical selection into the highest status. Collins and Zebrowitz's (1995) finding — that babyfacedness, height and weight may have a greater impact on status attainment in civilian than in military settings — may be due to this sampling effect.

Furthermore, height, which is correlated with status in some civilian populations (Frieze, Olson & Good 1990; Gillis 1982; Hensley & Cooper 1987), was purposefully dissociated from rank at West Point, where in the 1940s cadets were assigned to companies according to their height, some companies consisting of tall men and others of short men, a practice since discontinued. Because promotions at West Point are purposefully distributed nearly equally across the cadet companies, the effect in those days was to distribute promotions across heights as well. Once the cadets graduated, height was no longer explicitly constrained from being a factor in promotion, but some implicit

constraint may have persisted since the height of classmen did not correlate with their ranks in later career either (Mazur, Mazur & Keating 1984). Considering all these factors, we believe that the effects of physical features on promotion are relatively understated in our military population. We predict that civilian institutions such as church hierarchies will show stronger physical effects on promotion.

If physical features are important determinants of socioeconomic status in the general population, then their incorporation into sociological models of status attainment would add considerable power. Methodologically this is feasible, since we have shown that photographs taken decades years apart can be scored with similar reliability for facial dominance. Trained interviewers could photograph respondents as part of the interview.

Since physical features are heritable, they may explain some persistence in status from one generation to the next. If one's parents attained high (or low) status partially through their looks, then by inheriting a similar look, one ought to be biased toward similar status attainment. Possibly controlling for looks would substantially lessen the traditional correlation between the social status of parents and their children. In our sample, however, military background of father had no impact on career chances (see the discussion of results) and, therefore, can be excluded as the underlying cause of the positive effect of facial dominance.

We believe it likely that there are interactions involving looks, race, and status attainment. For example, a dominant-looking black male in a white majority society might suffer inordinate suppression because he looks threatening, whereas a submissive-looking black man, by looking less intimidating, might be trusted more in business or school. For whites, who are not subject to fearful discrimination, dominant looks would facilitate upward mobility. Adding physical appearance to traditional models of status attainment would open wholly new possibilities for explanation.

References

Andersen, Per Kragh, O. Borgan, R.D. Gill, and N. Keiding. 1992. *Statistical Models Based on Counting Processes*. Springer.

Atkinson, R. 1981. "Blood Brothers. The West Point Class of '66." *Kansas City Times*, 17 Nov.

_____. 1989. *The Long Gray Line*. Houghton Mifflin.

Barkow, Jerome H., L. Cosmides, and J. Tooby. 1992. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford University Press.

Bentler, P., and C. Chou. 1987. "Practical Issues in Structural Modeling." *Sociological Methods and Research* 16:78-118.

Berry, D.S. 1990. "Taking People at Face Value: Evidence for the Kernel of Truth Hypothesis." *Social Cognition* 8:343-61.

_____. 1991. "Accuracy in Social Perception: Contribution of Facial and Vocal Information." *Journal of Personality and Social Psychology* 61:298-307.

Berry, D.S., and L.A. Zebrowitz-McArthur. 1986. "Perceiving Character in Faces: The Impact of Age-Related Cranio-Facial Changes on Social Perception." *Psychological Bulletin* 100:3-18.

Berscheid, E., and E. Walster. 1974. "Physical Attractiveness." Pp. 157-215 in *Advances in Experimental Social Psychology*. Vol. 7, edited by Leonard Berkowitz. Academic Press.

Blau, Peter, and Otis D. Duncan. 1967. *American Occupational Structure*. Wiley.

Breslow, N. 1974. "Covariance Analysis of Censored Survival Data." *Biometrics* 30:89-99.

Butler, R.P. 1971a. "Official Reasons for Separations from Active Duty of USMA Graduates from the Classes of 1942 through 1966." Office of Institutional Research, U.S. Military Academy.

____. 1971b. "Survey of Careerists and Non-Careerists from the USMA Classes of 1963-1967." Office of Institutional Research, U.S. Military Academy.

____. 1976. "Relationships between College Performance and Success as an Army Officer." *Journal of Vocational Behavior* 9:385-91.

Child, Irwin. 1950. "The Relation of Somatotype to Self-Ratings on Sheldon's Temperamental Traits." *Journal of Personality* 18:440-53.

Collins, M.A., and L.A. Zebrowitz. 1995. "The Contributions of Appearance to Occupational Outcomes in Civilian and Military Settings." *Journal of Applied Social Psychology* 25:129-63.

Cortes, Juan, and Florence Gatti. 1972. *Delinquency and Crime*. Seminar.

Cox, David Roxbee. 1972. "Regression Models and Life-Tables." *Journal of the Royal Statistical Society* 34:187-220.

Cox, D.R., and V. Isham. 1980. *Point Processes*. London: Chapman & Hall.

Cox, D.R., and D. Oakes. 1984. *Analysis of Survival Data*. London: Chapman & Hall.

Enquist, M. 1985. "Communication during Aggressive Interactions with Particular Reference to Variations in Choice of Behaviour." *Animal Behaviour* 33:1152-61.

Fleming, Thomas R., and David P. Harrington. 1991. *Counting Processes and Survival Analysis*. Wiley.

Frieze, I.H., J.E. Olson, and D.C. Good. 1990. "Perceived and Actual Discrimination in the Salaries of Male and Female Managers." *Journal of Applied Social Psychology* 20:46-67.

Gillis, John. 1982. *Too Tall Too Small*. Institute for Personality and Ability Testing.

Gowin, Enoch Burton. 1915. *The Executive and His Control of Men*. Macmillan.

Grafen, A. 1990. "Biological Signals as Handicaps." *Journal of Theoretical Biology* 144:517-46.

Guthrie, Russell Dale. 1976. *Body Hot Spots*. Van Nostrand Reinhold.

Harper, D.G.C. 1991. "Communication." Pp. 374-96 in *Behavioural Ecology*, edited by J.R. Krebs and N.B. Davies. Oxford: Blackwell.

Hasson, O. 1994. "Cheating Signals." *Journal of Theoretical Biology* 167:223-38.

Hatfield, Elaine, and Susan Sprecher. 1986. *Mirror, Mirror: The Importance of Looks in Everyday Life*. SUNY Press.

Hensley, W.E., and R. Cooper. 1987. "Height and Occupational Success: A Review and a Critique." *Psychological Reports* 60:843-49.

The Howitzer. 1950. U.S. Military Academy.

Janowitz, Morris. 1971. *The Professional Soldier*. Free Press.

Keating, C., Allan Mazur, and M. Segall. 1981a. "A Cross-cultural Exploration of Physiognomic Traits of Dominance and Happiness." *Ethology and Sociobiology* 2:41-48.

____. 1981b. "Culture and the Perception of Social Dominance from Facial Expression." *Journal of Personality and Social Psychology* 40:615-26.

Lin, D.Y., and L.J. Wei. 1991. "Goodness-of-Fit Tests for the General Cox Regression Model." *Statistica Sinica* 1:1-17.

Lin, D.Y., L.J. Wei, and Z. Ying. 1993. "Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals." *Biometrika* 80:557-72.

Mazur, Allan. 1985. "A Biosocial Model of Status in Face-to-Face Primate Groups." *Social Forces* 64:377-402.

Mazur, Allan, J. Mazur, and C. Keating. 1984. "Military Rank Attainment of a West Point Class: Effects of Cadets' Physical Features." *American Journal of Sociology* 90:125-50.

Mazur, Allan, and Ulrich Mueller. 1993. "Persistence of Facial Dominance Measured from Portraits Taken Twenty Years Apart." Manuscript.

____. N.d. "Channel Modeling: From West Point Cadet to General." *Public Administration Review*. In press.

Mazur, Allan, and Leon Robertson. 1972. *Biology and Social Behavior*. Free Press.

Møller A.P. 1987. "Social Control of Deception among Status Signaling House Sparrows *Passer Domesticus*." *Behavioral Ecology and Sociobiology* 20:307-11.

Moore, D., and B. Trout. 1978. "Military Achievement: The Visibility Theory and Promotion." *American Political Science Review* 72:452-68.

Mueller, Ulrich. 1991. *Social and Reproductive Success: Theoretical Considerations and a Case Study of the West Point Class of 1950*. Mannheim, Ger.: ZUMA.

Mueller, Ulrich, and Allan Mazur. 1995. "Facial Dominance in Male *Homo Sapiens* as Honest Signaling." Manuscript.

Patzer, Gordon. 1985. *The Physical Attractiveness Phenomena*. Plenum Press.

Register of Graduates and Former Cadets. 1964, 1980, 1991. Association of Graduates, U.S. Military Academy.

Sheldon, William Herbert. 1942. *The Varieties of Temperament*. Harper & Row.

Smuts, Barbara B., D.L. Cheney, R.M. Seyfarth, R.W. Wrangham, and T.T. Struhsaker. 1987. *Primate Societies*. University of Chicago Press.

Stogdill, Ralph Melvin. 1974. *Handbook of Leadership*. Free Press.

Truscott, Lucian K., Jr. 1978. *Dress Gray*. Fawcett Crest.

Zahavi, A. 1977. "The Cost of Honesty." *Journal of Theoretical Biology* 67:603-5.

Zebrowitz, L.A., K. Kendall-Tackett, and J. Fafel. 1991. "The Influence of Children's Facial Maturity on Parental Expectations and Punishments." *Journal of Experimental Child Psychology* 52:221-38.

Zebrowitz, L.A., and J. Montepare. 1992. "Impressions of Babyfacedness and Attractiveness across the Life Span." *Developmental Psychology* 28:1143-52.

Zebrowitz, L.A., K. Olson, and K. Hoffman. 1993. "Stability of Babyfacedness and Attractiveness across the Life Span." *Journal of Personality and Social Psychology* 64:453-66.

Minority Access to White Suburbs: A Multiregional Comparison*

JOHN R. LOGAN, *State University of New York at Albany*
RICHARD D. ALBA, *State University of New York at Albany*
SHU-YIN LEUNG, *State University of New York at Albany*

Abstract

What are the locational processes that underlie racial residential segregation? This study analyzes the residential patterns of non-Hispanic whites, blacks, Hispanics, and Asians in eleven suburban regions using a multilevel model of locational attainment: the racial composition of a person's place of residence is determined by the person's individual characteristics (e.g., socioeconomic status and cultural assimilation) and aggregate variables including his or her racial/ethnic group, characteristics of the region in which he or she lives, and average characteristics of the group in that region. With some exceptions, findings for individual effects are more consistent with assimilation theory for Hispanics and Asians and with racial stratification theory for blacks. Having controlled for individual effects, disparities with whites are greatest for blacks, and for all three minority groups they are greatest in regions with large minority populations.

Residential segregation is one of the most enduring topics in the sociological literature on race and ethnicity (Massey & Denton 1993), but despite the decades-long effort devoted to its study, considerable uncertainty persists over the sources of observable levels of segregation. In part, this is because racial and ethnic residential segregation occurs in the context of a metropolis that is also segregated by social class and family life cycle (Shevky & Bell 1955). Since racial and ethnic groups typically differ in class and family characteristics, it is not obvious to what extent observed group differences in settlement patterns should be attributed to race and ethnicity and to what extent to other sources of segregation. In addition, sociologists have largely approached the study of segregation through the analysis of aggregate indices, such as the well-known index of dissimilarity (e.g., Lieberson 1963; Massey & Denton 1987). Problems of ecological inference have thus added to the difficulty of accurately assessing

* *A preliminary version of this article was presented at the 1992 annual meeting of the Population Association of America. This research was supported by grants from National Science Foundation (SES-8921273) and the National Institute of Child Health and Human Development (R01 HD25250). Direct correspondence to John Logan, Department of Sociology, State University of New York, Albany NY 12222.*

the respective roles of the various personal and household characteristics that help to determine residential location. Our main purpose here is to estimate models at both the individual and aggregate levels (a form of multilevel analysis). This allows us (1) to disentangle the effects of race or ethnicity from associated characteristics such as socioeconomic status and life cycle, and (2) to distinguish between regional effects (that is, processes that affect all group members in a region) and individual effects (that is, processes that depend on a person's own attributes).

A Multilevel Approach to Racial and Ethnic Segregation

Early work by Duncan and Duncan (1955) and Taeuber and Taeuber (1965) established the utility of comparing measures of black-white segregation across metropolitan regions and over time. A handful of studies (such as Lieberson & Carter 1982; Logan & Schneider 1984; Van Valey, Roof & Wilcox 1977) reported segregation measures for city and suburban areas in the 1960-1980 period. These studies provided some indications of trends in segregation and the differences between the residential experiences of blacks and those of other minorities.

The field took a large step forward with the research program organized by Douglas Massey at Chicago in the 1980s (culminating in the recent publication of Massey & Denton 1993). Massey's group systematically combed census files for 1970 and 1980 to provide multiple measures of segregation (usually reporting on the index of dissimilarity and P* exposure measures) for blacks, Hispanics, and Asians, and for both the city and suburban portions of sixty metropolitan regions. Their essential strategy was to model differences in levels of segregation, estimating the degree to which segregation covaries with average characteristics (e.g., the median socioeconomic level) of group members.

Such analyses are inherently limited. To the extent that they are intended to examine true contextual effects (such as the effects of minority group size or housing market characteristics or age of the housing stock on segregation) — that is, effects that are presumed to affect all group members in the metropolitan region — their conclusions may be soundly based. But the step that one cannot take in a convincing way with segregation indices is to examine the individual determinants of location that produce the patterns of segregation observed at an aggregate level. As noted by Hauser (1974), apparent aggregate effects may mask any combination of individual and aggregate causal processes; inference from aggregate data to individual effects can be made without bias only where there are no real aggregate causal processes (Firebaugh 1978). This concern is also the basis for the multilevel models proposed by Mason, Wong, and Entwisle (1983).

The problematic nature of inferences about individuals from aggregate analyses is easily illustrated. A central question for research has been the effect of income on segregation, because it addresses the hypothesis that socioeconomic achievement of minority group members translates into spatial locations that more closely resemble those of the majority. Two analyses, both reported in Massey and Denton (1987) and both based on aggregate data, are especially

relevant to this hypothesis. In both cases, their analyses led to conclusions very different from those conducted for individuals.

In the first analysis, the dependent variable is the proportion of blacks (or Hispanics or Asians) living in the suburban area of the metropolitan region (taken as an indicator of breaking the spatial isolation of the inner city). The key predictor, income, is operationalized as the median family income of group members in the region. For no group is the coefficient for income significant at the .05 level. In contrast, Alba and Logan (1991) analyze the same question with an individual model (where the city versus suburban location of individuals is predicted by their own household income, among other variables). At the individual level, there are strong effects of income for all groups, though less strong for blacks than for other minorities.

In their second analysis, Massey and Denton predict the level of segregation of metropolitan regions. Segregation is indexed by the P* measure of "contact with Anglos," which is equivalent to the proportion that is non-Hispanic white in the place where the average minority group member lives. Again the median family income of group members in the region (tested in separate equations for blacks, Hispanics, and Asians) is included as a predictor. Surprisingly, a negative effect of income is found for Hispanics, with no significant effect for blacks or Asians. In sharp contrast, Alba and Logan (1993) report parallel analyses at the individual level for one suburban region (where the dependent variable is the proportion that is non-Hispanic white in the community where the person resides and the person's household income is included as a predictor). There are strong and significant positive effects for all three minority groups in New York suburbs.

If we interpret these apparently disparate results carefully, there is no contradiction between them. This is because sociologists are interested in two kinds of theoretical questions. Some questions call for comparisons among a group's experiences in different parts of the country (e.g., is the housing discrimination faced by a minority group's members less severe in regions where the group has a higher socioeconomic status?). Others clearly refer to processes that occur at the individual level (e.g., do more affluent members of a minority group in a given region have greater proximity to whites?). In principle, it could be true that *a person's* income has a strong positive effect on individual outcomes, but that the average income of *a group* has no independent effect.

On what empirical grounds can we reach conclusions about causal processes at each level? Our main point here is that cross-regional comparisons of segregation indices may be inappropriate at both levels. First, what is true of group averages may not be true for individuals: this is the point of the ecological fallacy. Second, even where the theoretical question is posed at the group level, cross-regional comparisons can be misleading if they do not control for individual processes. This is the point of much of the debate over contextual effects, where some have argued that what seemed to be contextual effects might disappear if individual effects were adequately controlled. Standard segregation indices include no such controls. But as we shall demonstrate, there is an alternative approach, in which segregation is studied first at the individual

level within regions and then measures of white-minority differences in location — measures that adjust for individual effects — are compared across regions.

Theories of Minority Residential Location

Before we describe our procedure, we can make the discussion more concrete by introducing the general theoretical models that are implicated in segregation research and their implications for hypotheses at both levels. The main theoretical debates can be organized around two broad approaches. From one influential perspective on race relations in the U.S., segregation mostly reflects group differences in resources and preferences. This perspective emerged out of the experience of European immigrant groups over the last century (Lieberson 1963; Lieberson & Waters 1988; Massey 1985). As minorities become assimilated into mainstream society, participating in American culture and advancing in education and class status, they are expected to move out of traditional enclaves and ghettos. Achievement, not ascription, is the key to opportunity, in the housing market as well as in factory and field. Massey (1985) has described this model with the term "spatial assimilation."

The persistent and high level of black-white segregation, by contrast, has supported a model of racial and residential stratification in American society. Observers have described a variety of mechanisms of public and private discrimination (see Farley & Allen 1987; Galster 1988). These mechanisms include both individual and institutional actions, ranging from acts of violence against the pioneer minorities in a community to restrictive zoning (Logan & Stearns 1981; Rieder 1985; Shlay & Rossi 1981). It has been argued that a dual housing market operates for blacks (and perhaps for other minorities). This means, for example, that realtors may show prospective white and minority home buyers houses in different neighborhoods and that financial institutions are more cautious in granting mortgages to minority buyers (for an extended discussion of discriminatory mechanisms, see Massey & Denton 1993). These generate racially segmented housing markets, which make it difficult for minorities to enter some communities and add to the cost they must pay for housing (Berry 1975; Foley 1973; Stearns & Logan 1986). From this perspective, racial inequality is an integral part of the social structure and is reflected in the unequal spatial distribution of groups. We refer to this as a place stratification model, because it is derived from a theory of how political processes impinge on community development (Logan 1978; Logan & Molotch 1987). The aggregate effects of the dual housing market are to reinforce inequalities between places. It might also be termed a racial stratification model.

These perspectives can be viewed as overlapping as much as they are competing. Where they are most clearly incompatible is in their assessment of black-white differences. From one side, the black case is viewed as exceptional (Massey & Denton 1993), or — for some observers, judging from the progress believed to have been made by the black middle class — perhaps in transition (Farley & Frey 1994). From the other side, it is the success of white ethnic groups that is considered exceptional, while emphasis is placed on the continuity of discrimination faced even by the most affluent blacks.

However, the application of these perspectives to the experiences of Asian and Hispanic groups is much less developed, despite the explosive growth of these minorities as a result of recent immigration. Thus, Asian and Hispanic groups offer new possibilities for testing aspects of the assimilation and stratification models. How are these new groups, some of whose members are relatively well educated and wealthy, being incorporated into American metropolitan areas? Is their settlement pattern consistent with their personal backgrounds, and can they achieve residential integration with the non-Hispanic white majority? Much research on these groups has examined their creation of ethnic economies (Wilson & Portes 1980; Zhou 1992). It has been argued that there is a high degree of voluntarism — and success — associated with the economic and residential boundaries that separate them from mainstream society (for the same argument applied to blacks, see Schelling 1971 and the recent investigation by Farley et al. 1994). What are the implications of such distinctive strategies for locational patterns of group members? Do assimilation and stratification models now have to be supplemented with an "enclave model" of residential incorporation? In particular, is there evidence that more affluent persons among these immigrant minorities are choosing not to live in "whiter" suburbs?

A number of hypotheses will direct our analysis. As noted, the spatial assimilation model indicates that the residential situation of minorities, including their degree of integration into the majority, changes with social mobility and assimilation. This notion leads to two broad hypotheses about individual effects:

Hypothesis 1. For the members of any minority group, residential integration with whites increases as their socioeconomic status increases.

Hypothesis 2. For the members of any minority group, residential integration increases with cultural assimilation.

Whites, regardless of their personal and household characteristics, generally reside in largely white communities. Nevertheless, socioeconomic status and assimilation should also affect the kind of suburb in which non-Hispanic whites live. As noted above, research with aggregate indices has yielded conflicting results on these hypotheses. Only a handful of studies have been conducted at the individual level. Gross and Massey (1991), Massey and Denton (1985), and Villemez (1980) found positive effects of socioeconomic status for whites, blacks, and (in the two later studies) Hispanics on the proportion of the population that was white in census tracts. Alba and Logan (1993) found positive effects of socioeconomic status for whites, blacks, Hispanics, and Asians for suburbs in the New York region. Assimilation measures, however, such as nativity and language, had no effect for Asians, negative effects for blacks, small positive effects for whites, and strong positive effects for Hispanics.

Researchers in both the assimilation and stratification traditions agree that African Americans suffer the most extreme segregation and also that this segregation is not explained by income differences between the white and black groups (Farley & Allen 1987). Hypothesis 3 extends this finding to a broader set of relevant characteristics.

Hypothesis 3. At no level of socioeconomic and assimilation status do African Americans achieve residential integration with whites on a par with other minorities (or whites). In operational terms, the predicted percentage of non-Hispanic whites in the community of residence of African Americans (taking into account relevant personal characteristics) is lower than in communities of similar members of other groups, majority and minority.

Research at both the aggregate level (e.g., Massey & Denton 1987) and the individual level (Alba & Logan 1993; Gross & Massey 1991; Massey & Denton 1985; Villemez 1980) reveals two relevant results. First, net differences between blacks and other groups remain after controlling for other variables. Second, the coefficients linking socioeconomic status to locational outcomes are typically lower for blacks than for other groups.

Predictions for other minorities are more ambiguous. From the perspective of the stratification model, these groups are distinctive minorities whom whites frequently attempt to exclude from their communities, though less systematically than is true of African Americans. From the perspective of the assimilation model, Asians and Hispanics can be seen as immigrant groups (Puerto Ricans excepted) subject to the processes of assimilation. Therefore they should achieve residential parity with whites once they attain favorable socioeconomic and assimilation characteristics. This is the finding of Alba and Logan (1993) for the New York suburbs, supporting hypothesis 4:

Hypothesis 4. At higher levels of socioeconomic and assimilation status, Asians and Hispanics are residentially integrated with whites to the same degree as whites with similar characteristics.

Hypotheses 1-4 are stated for individuals and must be tested with individual-level models. Other hypotheses deal specifically with effects that can be tested only through interregional comparisons. We begin with the aggregate or context-level counterpart of hypotheses 1 and 2. Irrespective of a person's own attributes, processes of segregation may be ameliorated for a group as a whole insofar as it resembles the majority population in its socioeconomic or assimilation characteristics.

Hypothesis 5: The segregation of a minority is reduced where its average socioeconomic or assimilation status is higher.

Fundamental to the place-stratification model is the notion that whites seek to avoid residence with certain minorities. It is plausible that this avoidance would be more pronounced where these minorities are large in number (and, inversely, that whites would be less likely to erect residential barriers against small numbers of minority-group residents). In addition, where large numbers of a minority group are present, they may form their own residential enclaves. This reasoning leads to a region-level hypothesis concerning minority-group size, similar to the suppositions of other researchers (e.g., Blalock 1967; Massey 1985; Stearns & Logan 1986):

Hypothesis 6. Because whites tend to segregate themselves from large concentrations of minorities and because large groups may form their own suburban residential enclaves, the degree of segregation of a minority group should vary with its size. The effect should

be largest for African Africans, because white attempts to preserve distance from them are most systematic.

One common measure of segregation, the P* index of exposure between groups, depends inherently on group size. All else equal, a larger minority group has a lower exposure to whites, but the exposure of whites to the minority group is higher. We must be careful, therefore, to operationalize segregation in a way that avoids such bias.

Finally, racial/ethnic stratification across communities is part of a more complex system of stratification among communities, in which social class also plays a salient role. We expect that where communities are more stratified in general, there is more opportunity for racial/ethnic stratification by place. Therefore we hypothesize

Hypothesis 7. In regions where communities are more highly stratified (reflected in greater income disparity between suburbs and the central city, or among suburbs themselves), the segregation of racial/ethnic minorities tends also to be greater. This effect should be most pronounced for African Americans.

Research Design

In this article, we address these hypotheses through an investigation of minority residential integration with whites in suburbia, developed from a comparison of four major racial ethnic groups (non-Hispanic whites, non-Hispanic blacks, Hispanics, and Asians) in the eleven largest metropolitan areas in the nation in 1980. We employ a multilevel research approach. Assimilating and segregating processes *within* metropolitan regions are modeled at the individual level, including all relevant characteristics of minority-group members that may affect their access to communities with higher proportions of majority-group residents. Comparisons *across* regions, then, are based on segregation measures in which these individual processes have been controlled.

Such an approach is now possible, using readily available census data sets and by applying a methodology developed by Alba and Logan (1992-93; see also Alba, Logan & Bellair 1994 for a step-by-step review of the procedures). This method employs a model with the following form:

$$Y_{ij} = a + b_1X_{1ij} + b_2X_{2ij} + b_3X_{3ij} + \epsilon_{ij} \tag{1}$$

where Y_{ij} is a community-level measure for community *j* and thus is assumed to be constant for all individuals *i* in the same *j*; and the *X* variables are individual and household-level variables that predict residence in a community with a given level of the trait measured by the dependent variable. In the present case, Y_{ij} is a measure of community racial composition — specifically, the percentage of non-Hispanic whites among residents — which links it directly to a well-known measure of segregation, P* (Lieberson & Carter 1982; see Alba & Logan 1993 for a discussion of this linkage). The *X* variables cluster into certain types: X_{1ij} is the household income or another human-capital variable for individual or household *i* in community *j*, X_{2ij} is English-language ability or another assimilation measure, and X_{3ij} represents other (possibly

control) variables, such as household configuration or age. The model is estimated for the individual or the household as the unit and separately by racial or ethnic group.

Such a model can be interpreted in the same fashion as other status-attainment models (Blau & Duncan 1967). Each *b* coefficient establishes the degree of conversion of a minority individual's or household's characteristics, such as income, into a residential setting of low segregation (i.e., a community with many whites or, alternatively, nongroup members). Estimating effects of this sort has been a rather consistent goal of research on residential processes, but in the past it has been attainable only with specially constructed data sets or by an indirect route, using aggregate-level equations and thus risking the ecological fallacy (see, e.g., Massey, Condran & Denton 1987; Massey & Denton 1985; Villemez 1980; White, Biddlecom & Guo 1993).

To solve the problem of estimating locational attainment models at the individual level, we construct appropriate correlation (or covariance) matrices by combining correlations (or covariances) from two different kinds of U.S. census data sets. From the tabulations in Summary Tape Files (especially Summary Tape File 4B, which provides a wide variety of tables separately by racial or ethnic group), we estimate the correlations between the aggregate dependent variable and individual independent variables (these are the *XY* correlations). From Public Use Microdata Samples (PUMS), we estimate the correlations among the independent variables (the *XX* correlations). Ordinary least squares (OLS) regression equations are then estimated directly from the correlation matrices (along with means and standard deviations) prepared from these sources.

It may be easier to understand this procedure by imagining that we had begun with PUMS data to which somebody had appended the white percentage of each person's community of residence. Then we would have directly calculated the correlation between each individual characteristic and white percentage (the *XY* component of the matrix) and the correlations among predictors (the *XX* component of the matrix) from the same convenient data file. These would be exactly the same correlations, based on the same census sample data, as the ones that we calculate. Our method introduces no new conceptual complication.

Data and Variables

We study the suburban portions of the 11 largest metropolitan regions of the U.S. in 1980. When this analysis was completed in early 1995, the Bureau of the Census had only begun to release the group-specific Summary Tape Files from the 1990 census. Availability of those files will make it possible to test for changes in the patterns reported here. In order to avoid the sometimes arbitrary separations between neighboring primary metropolitan areas (e.g., between the New York PMSA and the Nassau-Suffolk PMSA), we have defined these metropolitan regions in the broadest possible fashion in each case, using the 1989 Consolidated Metropolitan Statistical Area (CMSA) boundaries.

The metropolitan areas in the analysis are displayed in Table 1, which also shows their constituent primary metropolitan areas and other data. The

suburban portions of these areas (1980 figures) encompass approximately 3.5 million Hispanics, 2.8 million non-Hispanic blacks, and 1.0 million Asians (along with 35.4 million non-Hispanic whites). The regions vary greatly in their minority presence, from Boston (whose suburbs are about 3.1% minority) to Los Angeles (where minorities comprise nearly 35% of the suburban population). They also vary in the predominant minority group. Hispanics are the largest minority in Los Angeles, Miami, San Francisco, Dallas, and Houston. Blacks are most highly represented in Washington and Miami. Asians are above 2% of the suburban population only in San Francisco, Los Angeles, and Washington.

Data suppression by the census, which does not report group-specific figures for places with few group members, reduces the number of places included in the following analyses. However, on average across the 11 metropolitan regions in this study, the included places encompass between 98% and 99% of non-Hispanic whites, non-Hispanic blacks, and Hispanics. The coverage of Asians averages 95%, ranging from 88% in Boston to 100% in Houston and Miami. The main effect of suppression is probably to understate the residential integration of suburban Asians, whom we will nevertheless show to be highly integrated with whites.

For some purposes, such as reflecting the sources of segregation between central city and suburban areas, analyses should include the entire metropolitan area. There are both theoretical and practical reasons for studying only suburban areas here (for analysis of city/suburb differences, see Alba & Logan 1991, Massey & Denton 1988, Stearns & Logan 1986). As Farley and Frey (1994:25) argue, the creation of segregated suburbs has been the key to persistence of segregation in the postwar era. Implicitly the opening of suburbs might be the key to its breakdown (see especially Frey & Speare 1988:240-41). For this reason, the suburban ring has drawn special attention from sociologists (Logan & Schneider 1984, Massey & Denton 1988, Stearns & Logan 1986).

Suburban communities are particularly relevant to the place stratification model. This model envisions communities as collective actors that employ a variety of strategies to protect or enhance their position in a hierarchy of places. Such communities are far more naturally operationalized as "places" in U.S. Census terminology than are the more familiar census tracts. For the most part, places in the census tabulations represent communities incorporated as villages, towns, or cities, although the Census Bureau also accepts as places locally recognized communities that have no separate legal existence. In such a scheme, however, it makes little sense to include central cities as places, given the different order of magnitude of their size. Finally, although analyses based on census tracts or block groups would reveal additional segregation *within* suburbs, the overall racial composition has significance in its own right, as expressed in such factors as the community's reputation and the composition of its school population.

In 1980 census data, there is an additional, practical reason to prefer places to tracts: namely, place-level tabulations for racial/ethnic groups (in Summary Tape File 4B) are based on more precise group definitions than their tract-level equivalents (in Summary Tape File 4A). In particular, separate tabulations for non-Hispanic whites and non-Hispanic blacks are available for places but not for tracts. At the tract level, tabulations for whites and blacks include Hispanics

TABLE 1: Metropolitan Area Definitions and Racial/Ethnic Composition of 1980 Suburban Population

CMSA *PMSA*	Percent White	Percent Black	Percent Hisp.	Percent Asian
Boston *Boston, Brockton, Lawrence-Haverhill, Lowell, Nashua, Salem-Gloucester* (339 places, mean population 12,789)	96.9	1.1	1.2	0.7
Chicago *Aurora-Elgin, Chicago, Gary-Hammond, Joliet, Lake County, Ill.* (457 places, mean population 9,811)	89.3	5.1	3.7	1.7
Dallas *Dallas, Fort Worth-Arlington* (208 places, mean population 7,890)	89.7	3.8	5.4	1.0
Detroit *Ann Arbor, Detroit* (272 places, mean population 12,652)	93.5	4.3	1.2	0.8
Houston *Brazoria, Galveston-Texas City, Houston* (150 places, mean population 9,352)	80.9	6.8	10.7	1.4
Los Angeles *Anaheim-Santa Ana, Los Angeles-Long Beach Oxnard-Ventura, Riverside-San Bernardino* (303 places, mean population 23,020)	65.8	6.5	23.0	4.5
Miami *Fort Lauderdale-Hollywood-Pompano, Beach Miami-Hialeah* (158 places, mean population 12,781)	65.6	13.0	20.4	0.7
New York-New Jersey *Bergen-Passaic, Jersey City, Middlesex-Somerset-Hunterdon, Monmouth-Ocean, Nassau-Suffolk, New York Newark, Orange County* (674 places, mean population 12,378)	86.6	7.1	4.8	1.4
Philadelphia *Philadelphia, Trenton, Vineland-Millville-Bridgeton, Wilmington* (458 places, mean population 8,038)	89.5	8.0	1.6	0.7
San Francisco *Oakland, San Francisco, San Jose, Santa Cruz, Santa Rosa-Petaluma, Vallejo-Fairfield* (188 places, mean population 18,084)	77.4	4.8	11.1	6.4
Washington, D.C. *Washington, D.C.* (279 places, mean population 9,218)	77.5	16.2	3.0	3.1

and thus overlap with tabulations for the Hispanic group. Therefore our analyses are restricted to suburban communities only.[1] We have nevertheless replicated these analyses for census tracts in the New York metropolitan region, and in that case the pattern of group differences is very similar to the one reported here (Logan et al. 1992).

Locational attainment models are estimated separately for each of four major racial/ethnic populations: non-Hispanic whites, non-Hispanic blacks, Hispanics, and Asians. Recognizing the potential importance of racial/ethnic differentiation within these groups, we include in each model (aside from that for non-Hispanic blacks) a series of dummy variables to represent subgroups. The model for whites has dummy variables for six major categories of single ancestry: English, French, German, Irish, Italian, and Polish; the omitted category contains individuals with any other single ancestry or with mixed ancestry. The model for Asians distinguishes among the following groups: Asian Indian, Chinese, Filipino, Japanese, Korean, and Vietnamese; the omitted category contains all other Asians. Finally, the model for Hispanics contains two sets of dummy variables. One set distinguishes among various nationality groups — Cubans, Mexicans, Puerto Ricans, and other Hispanics (omitted category) — and the other set among different racial groupings — white, black, and other (omitted category).

Other characteristics of individuals in our models are also represented in dummy variable format. This representation reflects the categorical presentation of variables in census tabulations. A side benefit is that this choice creates an intuitively compelling interpretation for the intercept, usually a largely forgotten parameter in regression interpretation, which takes on the predicted value for the percentage of non-Hispanic whites in the community of residence of an individual who is in the omitted categories of all the independent variables. The other independent variables in our models follow.

INDIVIDUAL ASSIMILATION

The measures of individual assimilation included are English-language ability (in the categories those who speak only English at home, who speak English well, and who do not speak English well), along with nativity and immigration status (categories: U.S. born, arrived before 1975, and arrived 1975 or later). The omitted categories are the first named in each case. Puerto Ricans, as U.S. citizens, are coded as U.S. born. As stated in hypothesis 2, the spatial assimilation model would anticipate less desirable spatial outcomes, and thus less residential integration with non-Hispanic whites, for recent immigrants, and more desirable locations for those who speak English well.

INDIVIDUAL SOCIOECONOMIC STATUS

Socioeconomic variables are home ownership (2 categories, with renter the omitted category), household income (9 categories, from under $5,000 to $75,000 and above, with the lowest category omitted), and education (5 categories, from grammar school to college graduate, with the lowest omitted). As formulated in hypothesis 1, a central tenet of spatial assimilation theory is that, controlling for

other relevant characteristics, minority-group members who have achieved higher socioeconomic standing will gain access to desirable locations equal to those of majority-group members of the same status. The theory implies a positive effect of socioeconomic variables on proximity to the majority.

DEMOGRAPHIC COMPOSITION

The demographic variables include age and household composition. Age is divided into five categories: individuals younger than 5, age 6 to 17, age 18 to 24, and 65 and older, with those of prime working age (25-64) forming the omitted category. Household composition is divided into four categories: married-couple households, other family types, nonfamily households, and institutional settings, with nonfamily households the omitted category. Both variables are included as necessary controls, because they have previously been shown to be independently associated with community characteristics and with suburban residence. Their effects are widely understood to indicate a preference, or greater willingness, to invest in location for married couples with children. These variables are also implicated in the construction of others, necessitating their explicit appearance in the models.[2]

The foregoing are the individual-level variables in locational attainment models that are separately estimated by group and metropolitan region. The *dependent variable* for these models is the percentage of non-Hispanic whites in the community (place) of residence.

As stipulated in hypotheses 5-7, our analysis also extends to intermetropolitan comparisons, in which individual processes have been controlled. The dependent variable at the regional level is the difference between the proportion that is white as predicted for a *minority group's members* and the proportion that is white as predicted for *whites* who have the same socioeconomic and assimilation characteristics.[3] Predictors, following previous cross-regional studies of segregation, are average characteristics of the minority in the region and characteristics of the region itself. We investigate group size, group composition, and suburban stratification.

GROUP SIZE

The minority versus white disparity is expected to be correlated with the proportions a group composes of the metropolitan region and of its suburban portion. According to hypothesis 6, these correlations should be positive.

GROUP COMPOSITION

These variables reflect the average characteristics of a group's suburban residents in the region: the percentage who have median household incomes of $40,000 and higher; the percentage who are college educated; the percentage who are U.S. born; the percentage who speak only English at home.

SUBURBAN STRATIFICATION

Suburban stratification in a metropolitan region is measured by the income disparity between suburban and central-city household incomes (index of net difference) and the income differences between places within suburbia (the standard deviation of median household incomes). Note that these measures tap different aspects of stratification: the index of net difference measures the overall socioeconomic selectivity of a region's suburbs, while the standard deviation measures the inequalities between suburban places. According to hypothesis 7, segregation can be expected to be greater where income stratification is greater.

Results

PREDICTING INDIVIDUAL EFFECTS

The first task is to understand how members of different racial/ethnic groups are channeled into communities of various compositions according to their personal and household characteristics. Table 2 summarizes the prediction equations for the four racial/ethnic groups in all 11 metropolitan regions. Because of space constraints the table lists only six parameters from these equations; complete regression equations are available from the authors on request (for the New York equation, see Alba & Logan 1993). Many of the omitted coefficients (such as those for various categories of income) are statistically significant but are redundant because the effect of the variable as a whole is reflected in the coefficient for its most extreme category. Nonsignificant coefficients are omitted. "Owner" is the coefficient for home ownership (as opposed to renting). "High Income" is the coefficient for persons with household incomes of $75,000 and above. "College" compares persons with college education to those with less than high school. "No English" compares those who do not speak English well with people who speak only English at home. Finally, "New Migrant" compares a post-1974 immigrant to native-born persons.

The intercepts are adjusted in two ways. First, they represent the typical case of persons in married-couple families (i.e., we have added the coefficient for married-couple families to the estimated intercept). Second, they represent the region's unique distribution of group members among national origins. As noted above, national origin categories for whites, Hispanics, and Asians are included in the equations, though they are not listed in Table 2. Predicted values from the equations depend, for example, on whether we assume that a Hispanic person is Mexican, Cuban, Puerto Rican, or "other." It would be unreasonable to assume the same category for every region (for New York, a Puerto Rican is most typical, but this is not true for Miami or Los Angeles). Therefore we have adjusted the intercepts to reflect what is "typical" for each region (adding a weighted average of the national origin coefficients, based on the number of persons of each origin in the region).

We begin by noting that the explained variance in these equations tends to be low, especially for whites. Of course one would not expect explained variance to be in the high range found in ecological studies, which ignore the

TABLE 2: Selected Coefficients from Regression Equations Predicting the Percentage of Non-Hispanic Whites in Suburban Place of Residence, by Racial/Ethnic Group and Metropolitan Region (CMSA), 1980

	Intercept	Owner	High Income	College Educ.	Poor English	New Immigrant	R^2
CMSA group							
Boston							
White	96.0	1.3	-0.6	-0.6	-0.8	-2.2	.07
Black	89.1	4.1	ns	ns	ns	-2.8	.16
Hispanic	93.1	2.3	ns	ns	-2.1	ns	.20
Asian	90.7	2.8	ns	ns	ns	ns	.23
Chicago							
White	89.6	1.3	2.9	1.0	ns	ns	.03
Black	53.3	-13.6	17.6	16.9	ns	12.4	.11
Hispanic	83.9	-0.9	7.8	4.6	-5.7	-2.0	.12
Asian	87.3	1.9	3.7	ns	ns	ns	.09
Dallas							
White	88.3	0.4	2.1	1.2	ns	ns	.02
Black	79.2	-5.3	8.8	ns	ns	5.0	.07
Hispanic	85.0	ns	ns	2.8	-4.3	2.4	.10
Asian	89.3	ns	ns	ns	ns	ns	.07
Detroit							
White	90.4	2.2	2.6	ns	ns	ns	.04
Black	55.9	-3.4	23.5	11.2	ns	ns	.07
Hispanic	81.6	2.2	12.1	5.9	-8.7	ns	.20
Asian	83.6	3.9	4.6	3.4	ns	ns	.07
Houston							
White	80.4	1.6	3.3	2.1	ns	ns	.03
Black	69.3	-1.2	ns	-3.6	ns	ns	.11
Hispanic	72.8	ns	5.4	5.4	-3.8	1.8	.06
Asian	82.0	ns	9.1	-3.8	ns	ns	.06
Los Angeles							
White	66.2	ns	9.0	8.6	-5.3	ns	.04
Black	28.5	-7.1	21.1	11.2	ns	14.0	.07
Hispanic	51.8	ns	14.1	9.4	-13.1	-3.9	.12
Asian	57.3	ns	9.1	4.8	-6.6	3.0	.04

TABLE 2: Selected Coefficients from Regression Equations Predicting the Percentage of Non-Hispanic Whites in Suburban Place of Residence, by Racial/Ethnic Group and Metropolitan Region (CMSA), 1980

CMSA group	Intercept	Owner	High Income	College Educ.	Poor English	New Immigrant	R^2
Miami							
White	73.0	1.1	3.8	3.6	-5.8	ns	.03
Black	37.2	-4.9	ns	8.3	ns	15.4	.04
Hispanic	55.6	ns	12.8	8.3	-14.3	1.9	.15
Asian	62.7	ns	ns	ns	ns	ns	.06
New York-New Jersey							
White	84.0	5.8	ns	ns	-4.4	ns	.06
Black	56.9	ns	15.0	ns	ns	ns	.04
Hispanic	69.5	9.8	9.0	7.0	-13.4	-3.8	.29
Asian	69.8	7.8	10.3	4.2	1.8	ns	.12
Philadelphia							
White	87.7	ns	2.1	1.5	ns	ns	.02
Black	63.3	-6.4	14.7	7.3	ns	9.6	.07
Hispanic	77.0	-5.5	20.4	ns	-11.2	8.5	.45
Asian	80.6	1.7	ns	3.7	ns	ns	.12
San Francisco							
White	75.7	1.0	5.3	4.6	ns	ns	.03
Black	55.1	-7.3	10.7	7.8	ns	6.0	.09
Hispanic	72.8	-0.7	8.4	4.0	-5.8	-1.4	.09
Asian	72.6	-2.1	9.0	1.9	-4.1	ns	.10
Washington							
White	77.6	2.4	2.5	3.0	ns	ns	.04
Black	65.4	-3.9	ns	ns	ns	8.9	.03
Hispanic	75.0	2.3	7.8	2.2	ns	ns	.15
Asian	69.3	5.2	9.3	3.1	ns	ns	.12

internal heterogeneity of communities. (Remember that the individual is the unit of analysis here. The number of cases depends on population size rather than the number of suburban places; in each region it is the number of cases reported in PUMS, with an arbitrary maximum of 10,000.) But the low range of R^2 for whites from .02 in Philadelphia to .07 in Boston also reflects the uniformity in white settlement patterns: most whites live in largely white communities regardless of their personal attributes of age, income, ethnicity, etc. In several suburban regions, the explained variance for blacks and Asians is comparable to that for whites, although it is higher in other regions. (The range for blacks

is from .03 in Washington to .16 in Boston, while for Asians it goes from .04 in Los Angeles to .23 in Boston.) In contrast, we find that background characteristics have their strongest effects on Hispanic settlement. The explained variance ranges from .06 (Houston) to .45 (Philadelphia). Hispanics tend to have greater heterogeneity in background characteristics than either blacks or Asians, to live in a greater variety of suburbs, and to have stronger relationships between their own background characteristics and the racial composition of their place of residence.

Let us consider now the regression coefficients themselves. According to hypothesis 1, income, education, and home ownership are expected to have positive effects on residence in suburbs with higher white proportions. Our results consistently support this expectation for whites. Out of 33 coefficients in the white equations (one for income, one for education, and one for home ownership in every region), 26 are statistically significant and positive. For Hispanics, most coefficients for income and education are significant and positive (18 of 22), but there are several cases where the home ownership coefficient is either not significant (4 cases) or is actually negative (3 cases). The exceptions for Asians include 4 cases each in which income, education, or home ownership is not significant. Finally, for blacks there are 4 cases each in which income or education is not significant and 9 cases in which the home ownership effect is negative.

These exceptions should not be overemphasized. Either higher income or higher education is associated with living in "whiter" suburbs for all three minority groups in almost every metropolitan region. Boston is an exception for minorities as well as for whites. Washington and Houston are exceptions for blacks; Miami and Dallas are exceptions for Asians. Members of minority groups with household incomes of more than $75,000 and/or college education typically live in suburbs with 5% or even 10% more white residents, in contrast to the poorest or least educated minority group members. And in most cases, the combined effects of these two indicators are greater for minorities than for whites. (This reflects our statement in the theoretical section of this article that most whites live in suburbs with high proportions of white residents, regardless of their personal background.)

But the results for home ownership point in a different direction. According to the assimilation model the additional resources represented by home ownership should increase choice about where one lives. For whites, in fact, there is typically a small positive effect of home ownership, so that white homeowners live in suburbs whose white percentage is 1 or 2 percentage points above what is predicted for white renters. There are two exceptions, Philadelphia and Los Angeles, while the largest effect is found in New York (5.8 points). By contrast, home ownership in most cases has a negative effect for blacks. Thus, in virtually all the regions (Boston is the one exception), black homeowners are further behind comparable whites than are black renters. Ownership does not behave as just another indicator of socioeconomic status. We suggest that the reason may lie in the dual-housing markets posited by the stratification model. If the market for home purchase by blacks were more geographically restricted than the market for rentals, then black owners would be more likely to reside in black neighborhoods. Indeed there is support elsewhere in the

literature for this interpretation: Stearns and Logan (1986) found that black suburbanization is greater in regions with a higher proportion of rental housing, and black-white segregation is lower in such regions.

The effects of home ownership are not as clear-cut for Asians and Hispanics. In some cases, the effects are positive, although only in New York could they be called substantial. In a few cases, they are negative, but mostly small. The largest negative effect is found for Hispanics in Philadelphia (-5.5). Thus, neither the assimilation nor the stratification model seems to fit the effect of ownership among Asians and Hispanics. Perhaps there are countervailing effects, leading to no net impact. Or perhaps whether home ownership is an asset or a liability for residential integration with the majority depends on other characteristics of the region (presumably aspects of the regional housing market) that we cannot identify here.

The next pair of variables — language and nativity — serve as indicators of cultural assimilation, which is central to the broader assimilation model as applied to immigrants (reflected in hypothesis 2). Among whites, there is only one case where immigration status affects residence: Boston, where recent immigrants are predicted to live in suburbs with 2.2% fewer white residents. There are more cases where poor English language ability affects whites, yielding about a 5-point difference in Los Angeles, Miami, and New York. The small number of significant effects suggests that white ethnicity, as reflected in recent immigration or poor English, does not have much éffect on the racial composition of suburbs where whites settle. Not surprisingly, English language ability proves to be irrelevant for blacks, whose native language (even for immigrants) is normally English. In seven cases, however, immigration is significant and — quite the contrary to expectations from assimilation models — positive. In the suburbs of Chicago, Los Angeles, and Miami, black immigrants live in suburbs with a white concentration as much as 15 percentage points higher. It should be noted that immigration status is paralleled by an ethnic distinction among non-Hispanic blacks, for West Indians (including Haitians) comprise the majority of the foreign born and often strive to preserve some distance from native-born African Americans (Kasinitz 1992).

Among the major recent immigrant groups, Hispanics and Asians, we find two opposite patterns. Hispanics, in ten of the eleven regions, are likely to live in less-white suburbs if their English is poor. For Asians, there are only two regions, San Francisco and Los Angeles, both of which have large and long-standing Asian settlements, where limited English language ability reduces Asian integration with whites. In most cases, immigration status is also nonsignificant. Thus cultural assimilation usually seems to be irrelevant to suburban Asians' residential patterns.

The results from Table 2 can be summarized as follows. First, as anticipated from the assimilation model, persons of higher income or education in most regions live in suburbs with higher proportions of white residents. Advancing socioeconomic status promotes suburban racial integration for blacks, Hispanics, and Asians (although higher socioeconomic status for whites simultaneously abets white exclusivity). Home ownership must be distinguished from other socioeconomic indicators because of the effects of dual-housing markets. Ownership is not typically associated with substantially greater minority

exposure to whites. And for blacks, in particular, ownership is most often linked to living in suburbs with more minority residents. Finally, assimilation as reflected in language ability is sometimes associated with living in whiter suburbs for whites, and in nearly every case for Hispanics. But neither assimilation variable is significant for Asians, and recent black immigrants actually have an advantage over native-born blacks.

THE NET EFFECTS OF INDIVIDUAL PROCESSES

The regression models summarized in Table 2 focus on the individual processes that lead some group members to have greater exposure to whites than do others. We have to this point avoided the issue of how the net results of these processes for racial/ethnic integration vary across the metropolitan regions. For this we need, for each group and each metropolitan area, a summary measure of residential integration with whites, net of the personal and household characteristics of individual minority-group members. We use several measures that reflect the difference between the percentage of non-Hispanic whites where a white lives (or is predicted to live) and the percentage of non-Hispanic whites where a minority person lives. A larger value means a greater disparity; the few negative values are cases where the minority group member is actually predicted to live in a whiter community than is the corresponding white.

One measure is based on the *unadjusted* percentage of whites in the community in which the average group member lives (this is, in fact, the difference in traditionally calculated P* for exposure to whites as calculated for whites and for each other group). We also present four indices that adjust for individual characteristics. The difference between the *predicted value* for whites and members of each minority group is a measure of the extent to which minority group members are residentially segregated even after adjusting for compositional differences. A group is more subject to segregating processes in a given region, we would say, when there is a greater discrepancy between the kind of suburb where group members and the kind where *comparable* whites live. This discrepancy depends on the coefficients of the locational attainment equations and on whatever combination of personal characteristics is assumed. We calculate indices for four combinations. The first is for individuals whom we describe as "poor immigrants": persons aged 25-65, in married-couple family households, renters, with incomes less than $5,000, who attended grammar school only, do not speak English well, and immigrated after 1974. A second is for "poor natives": the settings are the same as for the previous case, except that these individuals are native-born and speak only English at home (these correspond to the intercepts from Table 2). The third and fourth vary the socioeconomic variables. "Average" natives have the average characteristics of non-Hispanic whites across these eleven regions: they are homeowners and high school graduates with income of $20,000-29,999. Finally, "affluent natives" are homeowners with incomes of $75,000 or more who have completed college.

The Appendix table lists these values separately for all eleven metropolitan regions. For simplicity, we focus on the averages across regions presented in Table 3. Without introducing any adjustments for personal background (the first row of Table 3), we find that the average black lives in a suburb that has a

white percentage more than 25 points below the suburb of the average white. The average Hispanic is considerably more residentially integrated with whites, with a gap only half that magnitude. And this gap is even smaller for the average Asian. These figures confirm previously reported group differences in segregation in suburbia.

Now consider the impact of controlling for people's background characteristics other than race or ethnicity. The results vary across groups. Among Asians, these controls have much less impact on where they live than on where comparable whites live. This is because Asians differ from whites primarily in the proportion of immigrants among them, but we have seen that immigrant Asians in most suburban regions are about as residentially integrated with whites as are U.S.-born Asians or as those who speak only English at home. Asians have socioeconomic characteristics similar to those of whites, and in most regions the combined effect of socioeconomic variables is similar for Asians and whites, so controlling for this important predictor has little effect on white-Asian residential disparities. Some exceptions are found in the Appendix table. In New York-New Jersey and Washington, D.C., for example, the Asian disparity from whites declines clearly with socioeconomic advancement. But in most regions such a pattern is either small or nonexistent.

Among Hispanics, in contrast, the effects of introducing individual control variables are substantial and in the direction anticipated by assimilation theory. The category of Hispanics with the largest disparity with comparable whites is poor immigrants, because the cultural assimilation variables (language, nativity) have powerful effects on Hispanic residential location (much stronger than they do on whites). On average, there is almost no difference in predicted white percentage for affluent native Hispanics and affluent native whites. (We refer to white Hispanics; the dummy variables for race in the Hispanic equations, not included in Table 2, reveal that black Hispanics live in suburbs that are about 10% less white than do white Hispanics.) The pattern in which both cultural assimilation and socioeconomic advancement reduce Hispanic-white disparities is almost uniform across regions. One of the clearest cases is New York-New Jersey, where poor immigrant Hispanics live in suburbs of quite different racial composition than do poor immigrant whites (a 26.6 point disparity, greater than the mean difference between groups before introducing control variables). But for poor natives, the gap is cut nearly in half, and at the highest levels of income and education the difference actually favors Hispanics.

Finally we consider the black case. For the category of people that we identify as "poor immigrants," the difference from whites with the same backgrounds is much reduced among blacks (from 26.7 down to 10.8). For the equally poor category of natives, the disparity with comparable whites is much larger (23.1 percentage points). This reflects the fact that black immigrants are more residentially integrated than are native blacks. The socioeconomic variables also operate in a unique way for blacks. "Average native" blacks have a greater disparity with comparable whites than do "poor natives." This is because the "average" socioeconomic category includes home ownership, which we found to have negative effects on residential integration of blacks. The home ownership effect outweighs the positive income/education effect. Only at the

TABLE 3: Average Difference in Predicted Racial Composition (Percent Non-Hispanic White) between Whites and Minority Group Members with Various Combinations of Personal Characteristics

	Non-Hispanic Blacks	Hispanics	Asians
Not adjusted[a]	26.7	13.3	5.0
Poor immigrant	10.8	14.3	4.0
Poor native	23.1	9.1	5.7
Average native	27.2	3.9	4.9
Affluent native	17.8	1.7	2.4

[a] Equivalent to the difference in P* (exposure to non-Hispanic whites) for whites and minority group members

highest levels of income and education do these aspects of socioeconomic standing counterbalance home ownership.

This sort of examination of Table 3 bears directly on hypotheses 3 and 4, which predict that Asian and nonblack Hispanic integration will resemble that of whites themselves at higher levels of socioeconomic status and assimilation, but that African Americans will lag behind. Only for the unusual case of immigrant blacks are black-white disparities comparable to disparities measured for Hispanics or Asians. And for nonimmigrants, even a college degree and income over $75,000 do not result in black parity with whites. For this category, the disparity for blacks in 17.8 points, while near parity is found across these eleven regions for Hispanics (1.7) and Asians (2.4).

INTERPRETING REGIONAL VARIATION

Finally, we address regional variation in segregation. Previous studies have treated unadjusted measures of segregation (the index of dissimilarity or exposure index) as the dependent variable to be explained by other regional characteristics. The new measures introduced in Table 3 and the Appendix have the advantage of having controlled for individual processes. Again we simplify the presentation, using only the "average native" values. These are the most relevant, we believe, because they reflect situations in which all people are assumed to have the characteristics typical of the white population.

Table 4 reports the correlations of these indices with various characteristics of regions and of groups within regions. Our first region-level hypothesis (hypothesis 5) concerns the ameliorating impact of the average characteristics of group members: where the average status of a group's members is high, the predisposition of the majority to exclude them (or of the minority to segregate itself) may be lower. The correlations in Table 4 offer mixed support for this proposition for Hispanics and Asians, but show that segregation for blacks is

not affected by their average status. In the case of Hispanics, the percentage who are native born and the percentage who speak English at home are negatively correlated with net segregation from whites (-.64 for both variables). For Asians, the percentage who are college educated is also negatively correlated with segregation (-.76), but the correlation with Asian median income is smaller and not significant.

Another expectation is that a minority's size should affect the degree of its segregation from whites (hypothesis 6). We tested first for the size of each minority group. The correlations are around .80 for Hispanics both for the proportion the group comprises of the CMSA's overall population and also for its share of the suburban population. The positive direction is the one implied by hypothesis 6, for it indicates that the difference in residential pattern between whites and a minority grows as the size of the minority increases. The correlations are smaller for blacks and Asians. We also calculated correlations for the size of all three minorities added together — perhaps segregating processes depend not so much on the size of any one group as on the presence of minorities altogether. These coefficients are significant for all three groups (.56 for blacks, .80 for Hispanics, and .86 for Asians).

There is also some indirect evidence that size affects Hispanic segregation. In the regressions summarized in Table 2, we included dummy variables to represent subgroups (these are not reported in the table). In the Hispanic equations for seven regions, being Mexican is associated with less exposure to whites, and these include the regions in Texas and California where Mexicans are highly concentrated. In two cases the Mexican coefficient is positive; these are New York and Miami, where Mexicans are a small portion of the Hispanic population. There is a similar pattern for Puerto Ricans, with six cases of negative effects including New York; the only positive coefficient is for Houston, with few Puerto Ricans. Finally, Cubans are found to live in less white suburbs in four regions, including Miami; the only positive Cuban coefficient is in San Francisco. It appears, then, that the subgroup that constitutes the local Hispanic majority tends to be most segregated from whites in suburban regions.[4]

Suburban income stratification is also expected to be related to racial/ethnic segregation (hypothesis 7), because class-based exclusion may be a means for affluent whites to restrict the number of suburbanizing minorities in their communities. Table 4 includes two measures of income stratification, and the results are contrary to the hypothesis. Only one variable has significant correlations, and these are in an unexpected direction (the city/suburb index of net difference is negatively correlated with Hispanic and Asian disparities with whites). These results suggest that the extent of class segregation between city and suburbs or among suburbs is not an indicator of processes of racial exclusion.

Given the small number of regions involved in this analysis, it would be inappropriate to extend the correlations to a full-fledged multivariate analysis. By applying the analysis to 11 regional cases for each social group, however, we can offer a more parsimonious test of the results (yielding 33 cases for analysis). Our procedure was to conduct a stepwise regression in which dummy variables are entered first to represent the Hispanic and Asian minorities (blacks are the

TABLE 4: Correlations between the Discrepancy in Predicted Values[a] of Suburban Racial Composition for Whites and Minority Group Members and Various Characteristics of the Groups and Regions

Predictors	White Value minus Value for: Blacks	Hispanics	Asians
Socioeconomic status			
Percent group with college education	-.13	-.13	-.76*
Percent group with income over $40,000	.14	-.22	-.36
Assimilation			
Percent group born in U.S.	.18	-.64*	.32
Percent group speak English at home	.32	-.64*	.27
Size			
Percent group members in CMSA	.41	.83*	.29
Percent group members in suburbs	.46	.85*	.32
Percent all minorities in suburbs	.56*	.80*	.86*
Income stratification			
Suburb/city income difference	-.03	-.57*	-.64*
Income inequality between suburbs	-.02	-.17	.06
(N = 11)			

[a] Predicted values are for persons with these median or modal characteristics: age 25-64, married couple household, native born, speaks only English, homeowner, income $20,000-29,999 and high school graduate.

* $p < .05$ (two-tailed test)

omitted category). Then we test separately the significance of each variable from Table 4. Finally, for every such significant variable, we test its interaction effects with the race variables.

These procedures lead to a final simple model:

$$(1)\ \textit{Disparity} = 17.23 - 21.65\textit{Asian} - 13.42\textit{Hispanic} + .55\%\textit{Minority}$$
$$(\textit{Adjusted}\,R^2 = .654)$$

We find significant effects of race. The only variables that add significantly to race are those representing group size in the suburbs, and of these the strongest effect is for the combined size of all three groups. No interactions with race are significant. According to the coefficients, the disparities between whites and both Asians and Hispanics are substantially smaller than disparities with blacks, even after all individual sources of variation in residential placement are controlled. And net segregation for all groups is greater in suburban regions

where minorities comprise a larger share (and consequently, whites comprise a smaller share) of the population.

The relationships summarized in this equation can also be illustrated in a scatter diagram. Figure 1 shows how the difference between white and minority values (differences in the predicted white percentage in the suburb in which an "average native" person lives) is associated with the total minority proportion (black, Hispanic, and Asian) in a region. Every point represents a region. Points shown as small boxes represent white-black disparities. It is clear from Figure 1 that these are in every case higher than white-Asian or white-Hispanic disparities (both shown as asterisks). There is a clear tendency for disparities to increase with the relative size of the total minority population in the suburbs. And although the regression line for blacks is above the line for Asians and Hispanics (combined here for simplicity), it has a similar positive slope for both groups.

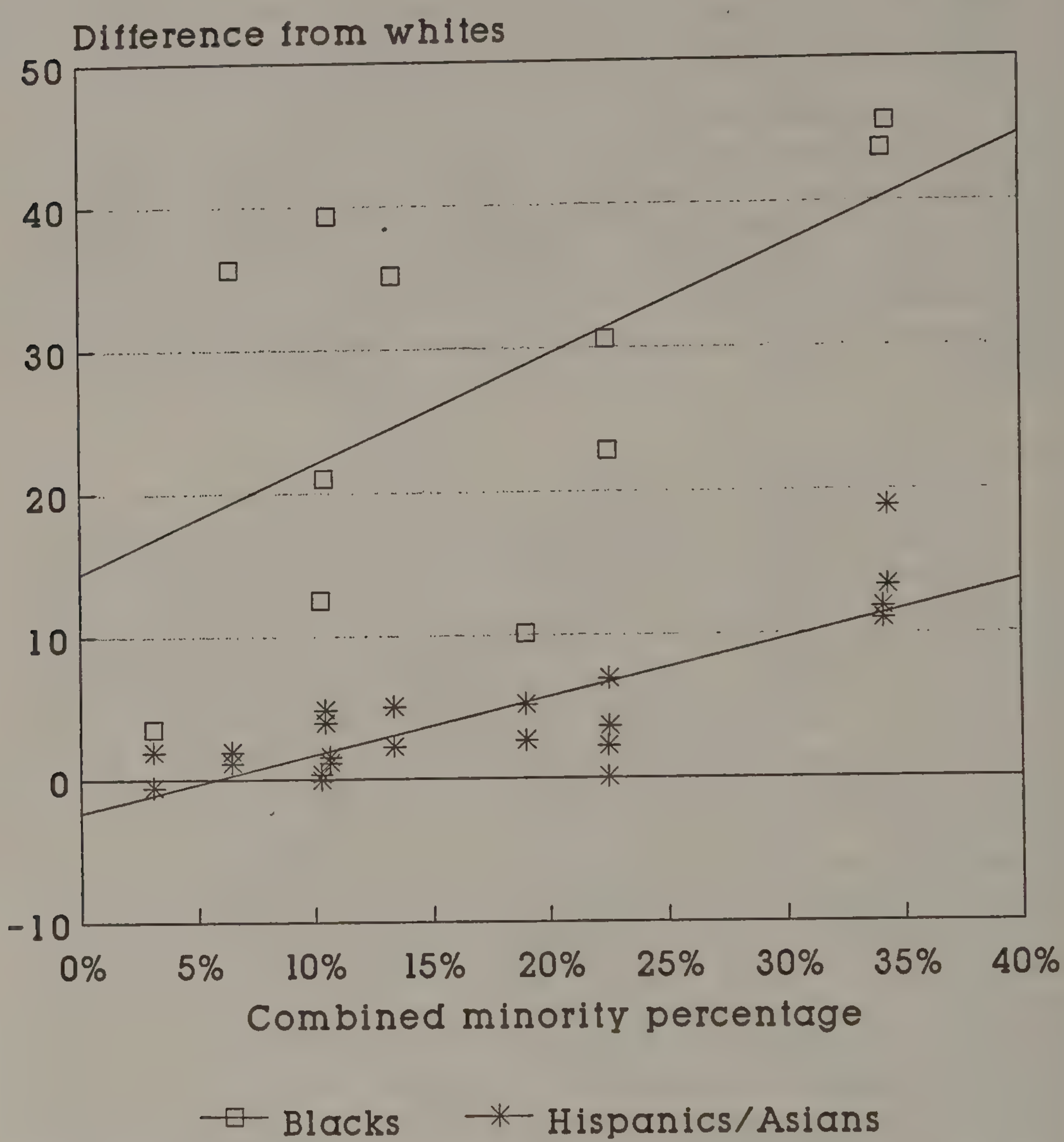
Summary and Conclusion

As others have pointed out, it is hazardous to draw inferences about processes affecting individuals from aggregate data, and it can also be risky to draw region-level conclusions from aggregate data without controlling for microlevel processes. Our approach is to model segregation (specifically, its inverse, coresidence with whites) in a multilevel fashion, disentangling the segregation processes occurring at the level of individuals (for a particular group and region) from those at the level of suburban regions (aspects of locational dynamics that affect all group members in the region). In this way, our procedure meets one of the long-standing goals of segregation research — to remove the effects of obvious determinants of location such as income from segregation measures (see, e.g., Farley & Allen 1987; Massey & Denton 1993).

We find, in brief, that several variables operate at the individual level to affect the racial composition of suburbs in which a person lives. These are not the same for every racial/ethnic group, but they include income, education, home ownership, nativity, and language. There is some indication that a few of these variables also have effects at the aggregate level in our sample. After controlling for individual processes, the zero-order correlations suggest that Asians are more integrated with whites in regions where the average education level of Asians is higher, and Hispanics are more integrated in regions where Hispanics are more assimilated. No such effects are found for blacks, however. For example, higher-income blacks live in whiter suburbs, but the difference between where a black person lives and where a comparable white person lives is not affected by the average level of black income in the region.

The one consistent regional effect is the effect of minority group size, which is found for all three groups in the correlations and also in the pooled regression equation. The clearest effect is that of the combined size of all three minorities. This is a true contextual effect: the disparity in residential location between minorities and comparable whites is greater in those regions with larger minority populations. We consider the evidence for this relationship to be

FIGURE 1: Difference in Predicted Percentage White for White vs. Minority Persons



strong. It should be pointed out, though, that different results might be found in a larger and more diverse sample of regions.

Aspects of both major theoretical models in the literature are supported in our findings. We do not assume that members of minority groups prefer to live in largely white areas. But segregation is a prime indicator of race relations: the assimilation model anticipates that segregation will decline as two groups become more similar in other respects, while the stratification model expects that segregation will prove more durable.

As the assimilation model asserts (hypothesis 1), higher income and education of individuals do result in greater residential exposure to whites for all minorities, blacks as well as Asians and Hispanics. As that model also

indicates (hypothesis 2), individual assimilation has some strong effects on proximity to whites, at least for Hispanics. However, such effects are missing for Asians, outside of two West Coast metropolitan regions where large Asian populations have established residential enclaves. This curious discrepancy between the major immigrant populations suggests differences in the process of residential segregation that deserve further conceptual and empirical investigation.

Additional support for the assimilation model comes from comparison of predicted values of racial composition for several standard types of people. The predicted values — in effect, net levels of segregation — indicate that the segregation of African Americans from whites is substantial, although somewhat smaller than unadjusted segregation scores would indicate. More relevant for the assimilation model is that Hispanics and Asians experience less net segregation than blacks, and those with the most favored background characteristics (high socioeconomic status for Asians, high socioeconomic status and good English-language ability for Hispanics) actually live in suburbs whose racial composition is similar to those of comparable whites (supporting hypothesis 4). Implied, of course, is that the segregation found for these groups in aggregate indices reflects the residential patterns for low-SES individuals and, in the case of Hispanics, for those who do not speak English fluently. There are some additional provisos concerning group differences within these two large populations. These assimilatory patterns do not apply to black Hispanics in general; and in particular regions, the largest Hispanic group for the region (e.g., Cubans in Miami) tends to be more segregated from whites than others.

The nearly complete residential integration of socioeconomically mobile Asians and Hispanics is contradictory to the notion of voluntary segregation of these groups. The most advantaged members, presumably, have the greatest choice of where to live. It appears that their preference — at least for those who live in the suburbs — is integration. If there is also a tendency of self-segregation, perhaps it is largely a central city phenomenon.

The stratification model finds support in several aspects of our results. The effects of a dual-housing market may be discerned in the negative impact of home ownership on blacks' residential integration with whites, a clear departure from the typically positive effects of other socioeconomic variables. Ownership is found to have mixed effects even for Hispanics and Asians, suggesting that the dual-housing market may have relevance for some new immigrant groups. On the regional plane, the positive relationship between group size and segregation from whites supports long-standing hypotheses emanating from researchers working within both theoretical frameworks. It is most consistent with the basic reasoning underlying the stratification model: that whites use segregation to preserve their social position in the face of a threatening — that is, large — minority advance (hypothesis 6).

Our results, it must be underscored, are limited to the suburban portions of the 11 largest consolidated metropolitan regions (CMSAs) in the country for 1980. The limitation of our analysis to suburbs cuts in two different ways. On the one hand, the selection of suburban communities as the geographic units for measuring segregation is justified by the notion, intrinsic to the stratification model, that communities are collective actors in the process of place stratifica-

tion. On the other, as Massey and Denton (1987, 1988) have argued, suburbanization is a key step in the process of residential integration for minorities, and the resulting greater integration with the majority should be reflected in less overall sorting of individuals into communities based on race and ethnicity and other characteristics. Thus, it is revealing that we have nevertheless found striking relationships between individual characteristics and the racial and ethnic compositions of the communities in which individuals reside.

These relationships may prove even stronger in analyses of location within the metropolitan area as a whole. Such analyses necessarily depend on using census tracts as the geographical units and represent one extension of the research program represented here. Another extension involves the metropolitan-area sample. The areas we have analyzed are not only the largest but, in general, the most racially and ethnically diverse. To the extent that the results of our models are consistent from region to region, conclusions from them can probably be safely extrapolated to other large and racially diverse areas. Nevertheless, it would be preferable to have a larger sample of regions, especially for the purpose of multivariate analysis across regions. And it is possible that results for smaller and less racially diverse areas may differ from those presented here. For both these extensions, our method of constructing locational attainment models provides a viable way of implementing the multilevel approach necessary for a convincing analysis of residential segregation.

Notes

1. Some boundary problems, minor for the most part, result from this choice. They stem from the fact that we must define the same geographic terrain in two different data sources — Summary Tape Files and Public Use Microdata Samples. The problems are not at the metropolitan level: since metropolitan regions and the geography of the PUMS data are both defined by county, it is not hard to make the metropolitan boundaries coincide fairly well in the two data sources. When discrepancies, typically involving small counties, arose, we generally dropped these counties from the analysis, thus forcing coincidence.

More significant divergences arise, however, because our study is focused on suburbs and thus central-city residents must be excluded. This is easily done in data from the Summary Tape Files, but not always in the PUMS. Small central cities (such as Sayreville, New Jersey, with a population under 30,000 in 1980) do occur in most of the metropolitan areas in our analysis and are not identifiable in the PUMS data we are using (A geography). Their residents are therefore included in our PUMS calculations, creating a small error in the sample definitions. The largest discrepancy occurs in Boston, where only Boston and Worcester can be identified and excluded in the PUMS analysis. Several smaller central cities with a total population of 665,000 are necessarily included (Brockton, Fall River, Haverhill, Lawrence, Lowell, New Bedford, Manchester, and Nashua). To evaluate the effect of this discrepancy, we reanalyzed this region as though these cities were suburban places (i.e., matching the PUMS and STF geography more exactly), and found substantially the same results.

2. Accordingly, interpretation of the effects of the demographic variables requires special care, and we will not focus on them. The crux of the matter is that some categories of both the age and household composition variables are necessitated by omissions from other variables (for additional discussion of this feature of the methodology as applied to the 1980 census, see Alba & Logan 1992-93). For instance, explicit representation of the group-quarters category is necessary in order to interpret properly variables such as household income, which are meaningful only for individuals in households. Unless a group-quarters dummy variable is included in the model, the omitted category for household income would implicitly mix the

group-quarters population into the lowest-income category, obviously muddying the interpretation. The problem cannot be solved by eliminating the group-quarters population, because the analysis depends on tabulations (e.g., for nativity/year of immigration) in which individuals in households and those in group quarters are inseparably mixed.

In like fashion, age categories accounting for the population younger than 25 are required by the education variable, which is defined only for persons 25 and older. The youngest age category is also required for the language variable, which has meaning only for persons aged 5 and older.

3. We consider several different "profiles" in order to summarize the net effects of variations in socioeconomic and assimilation status. For instance, in analyzing the regional determinants of segregation levels (Table 4), predicted values are for persons with the median or modal characteristics age 25-64, married couple household, native born, speaks only English, homeowner, income $20,000-29,999 and high school graduate.

4. Our equations for whites included dummy variables representing such ethnic subgroups as single-ancestry French, German, Italian, and other national origins. In some regions these variables have significant coefficients, but these are small and inconsistent (e.g., Italians tend to live in whiter suburbs in Detroit and Miami but less white suburbs in Washington). In line with the results for cultural assimilation variables, this suggests again that white ethnicity is not an important factor. For Asians, our equations included dummy variables representing Japanese, Korean, Chinese, Filipino, Indian, and Vietnamese national origins. The largest number of significant positive effects is for Japanese (positive in 7 regions, negative in 1), Koreans (positive in 6 regions, negative in 1), and Chinese (positive in 5 regions, negative in 2). The largest number of significant negative effects is for Filipinos (4 negative, 2 positive). This set of results suggests that there may be a hierarchy of Asian groups in suburbia, although clearly it is not uniform across regions.

References

Alba, Richard D., and John R. Logan. 1991. "Variations on Two Themes: Racial and Ethnic Patterns in the Attainment of Suburban Residence." *Demography* 28:431-53.

____. 1992-93. "Analyzing Locational Attainments: Constructing Individual-Regressions Using Aggregate Data." *Sociological Methods and Research* 20:367-97.

____. 1993. "Minority Proximity to Whites in Suburbia: An Individual-Level Analysis of Segregation." *American Journal of Sociology* 98:1388-1427.

Alba, Richard, John R. Logan, and Paul Bellair. 1994. "Living with Crime: The Implications of Racial and Ethnic Differences in Suburban Location." *Social Forces* 73:395-434.

Berry, Brian. 1975. "Short Term Housing Cycles in a Dualist Metropolis." Pp. 165-82 in *The Social Economy of Cities, Urban Affairs Annual Reviews*, no. 9, edited by Gary Goppert and Harold Rose. Sage.

Blalock, Hubert M., Jr. 1967. *Toward a Theory of Minority-Group Relations*. Wiley.

Blau, Peter, and Otis Dudley Duncan. 1967. *The American Occupational Structure*. Free Press.

Duncan, Otis Dudley, and Beverly Duncan. 1955. "A Methodological Analysis of Segregation Measures." *American Sociological Review* 20:210-17.

Farley, Reynolds, and Walter Allen. 1987. *The Color Line and the Quality of Life in America*. Oxford University Press.

Farley, Reynolds, and William Frey. 1994. "Changes in the Segregation of Whites from Blacks in the 1980s." *American Sociological Review* 59:23-45.

Farley, Reynolds, Charlotte Steeh, Maria Krysan, Tara Jackson, and Keith Reeves. 1994. "Stereotypes and Segregation: Neighborhoods in the Detroit Area." *American Journal of Sociology* 100:750-80.

APPENDIX: Differences in the Predicted Values of Percentage of Non-Hispanic Whites by Racial/Ethnic Group and Metropolitan Region (CMSA) for Selected Types of Group Members

	Non-Hispanic Black	Hispanic	Asian
Boston			
Unadjusted	4.6	3.9	3.3
Poor immigrant	8.3	2.7	2.3
Poor native	6.9	2.9	5.3
Average native	3.5	-.6	1.9
Affluent native	.4	.5	1.2
Chicago			
Unadjusted	35.8	9.0	2.0
Poor immigrant	18.8	13.9	3.4
Poor native	36.3	5.7	2.3
Average native	39.4	1.1	1.6
Affluent native	20.6	-.7	.8
Dallas			
Unadjusted	11.1	4.3	0.3
Poor immigrant	.2	4.7	-1.5
Poor native	9.1	3.3	-1.0
Average native	12.5	0	.3
Affluent native	7.2	1.9	.7
Detroit			
Unadjusted	34.7	6.4	2.0
Poor immigrant	17.7	14.2	2.1
Poor native	34.5	8.8	6.8
Average native	35.7	1.1	2.0
Affluent native	8.7	-2.4	.4
Houston			
Unadjusted	18.2	8.4	4.5
Poor immigrant	6.5	8.3	-3.1
Poor native	11.1	7.6	-1.6
Average native	10.7	2.6	5.2
Affluent native	16.7	3.5	2.0
Los Angeles			
Unadjusted	41.4	27.7	13.5
Poor immigrant	14.9	28.4	9.4
Poor native	37.7	14.4	8.9
Average native	43.7	11.0	11.9
Affluent native	30.9	8.9	13.3

APPENDIX: Differences in the Predicted Values of Percentage of Non-Hispanic Whites by Racial/Ethnic Group and Metropolitan Region (CMSA) for Selected Types of Group Members (Continued)

	Non-Hispanic Black	Hispanic	Asian
Miami			
Unadjusted	43.5	31.4	10.2
Poor immigrant	12.4	25.0	7.6
Poor native	35.8	17.4	10.3
Average native	45.5	18.8	13.3
Affluent native	39.3	3.9	2.7
New York			
Unadjusted	31.9	19.5	4.3
Poor immigrant	23.0	26.6	7.7
Poor native	27.1	14.5	14.2
Average native	35.1	2.2	5.0
Affluent native	18.7	-3.5	-.4
Philadelphia			
Unadjusted	23.8	22.1	2.0
Poor immigrant	11.0	23.5	3.9
Poor native	24.4	20.7	7.1
Average native	21.1	3.9	4.8
Affluent native	13.1	7.2	2.6
San Francisco			
Unadjusted	21.4	9.9	9.7
Poor immigrant	4.2	7.4	5.0
Poor native	20.6	2.9	3.1
Average native	22.8	3.6	6.9
Affluent native	20.3	2.0	5.2
Washington			
Unadjusted	27.8	4.0	4.1
Poor immigrant	3.0	3.2	7.7
Poor native	12.2	2.6	8.3
Average native	30.6	0.0	2.2
Affluent native	21.2	-1.8	-1.3

Firebaugh, Glenn. 1978. "A Rule for Inferring Individual-Level Relationships from Aggregate Data." *American Sociological Review* 43:557-72.

Foley, Donald. 1973. "Institutional and Contextual Factors Affecting the Housing Choices of Minority Residents." Pp. 85-147 in *Segregation in Residential Areas*, edited by Amos Hawley and Vincent Rock. National Academy of Sciences.

Frey, William, and Alden Speare. 1988. *Regional and Metropolitan Growth and Decline in the U.S.* Russell Sage.

Galster, George. 1988. "Residential Segregation in American Cities: A Contrary Review." *Population Research and Policy Review* 7:93-112.

Gross, Andrew, and Douglas Massey. 1991. "Spatial Assimilation Models: A Micro-Macro Comparison." *Social Science Quarterly* 72:347-60.

Hauser, Robert. 1974. "Contextual Analysis Revisited." *Sociological Methods and Research* 2:365-75.

Kasinitz, Philip. 1992. *Caribbean New York: Black Immigrants and the Politics of Race.* Cornell University Press.

Lieberson, Stanley. 1963. *Ethnic Patterns in American Cities.* Free Press.

Lieberson, Stanley, and Donna Carter. 1982. "Temporal Change and Urban Differences in Residential Segregation: A Reconsideration." *American Journal of Sociology* 88:296-310.

Lieberson, Stanley, and Mary Waters. 1988. *From Many Strands: Ethnic and Racial Groups in Contemporary America.* Russell Sage Foundation.

Logan, John R. 1978. "Growth, Politics, and the Stratification of Places." *American Journal of Sociology* 84:404-16.

Logan, John R., Richard D. Alba, Thomas McNulty, and Brian Fischer. 1992. "Making a Place in the Metropolis." Unpublished manuscript.

Logan, John R., and Harvey Molotch. 1987. *Urban Fortunes: The Political Economy of Place.* University of California Press.

Logan, John, and Mark Schneider. 1984. "Racial Segregation and Racial Change in American Suburbs, 1970-1980." *American Journal of Sociology* 89:874-88.

Logan, John, and Linda Stearns. 1981. "Suburban Racial Segregation as a Nonecological Process." *Social Forces* 60:61-73.

Mason, William, G. Wong, and Barbara Entwisle. 1983. "Contextual Analysis through the Multilevel Linear Model." Pp. 72-103 in *Sociological Methodology 1983-1984*, edited by Samuel Leinhardt. Jossey-Bass.

Massey, Douglas. 1985. "Ethnic Residential Segregation: A Theoretical Synthesis and Empirical Review." *Sociology and Social Research* 69:315-50.

Massey, Douglas, Gretchen Condran, and Nancy Denton. 1987. "The Effect of Residential Segregation on Black Social and Economic Well-Being." *Social Forces* 66:29-56.

Massey, Douglas, and Nancy Denton. 1985. "Spatial Assimilation as a Socioeconomic Outcome." *American Sociological Review* 50:94-106.

———. 1987. "Trends in the Residential Segregation of Blacks, Hispanics, and Asians: 1970-1980." *American Sociological Review* 52:802-25.

———. 1988. "Suburbanization and Segregation in U.S. Metropolitan Areas." *American Journal of Sociology* 94:592-626.

———. 1993. *American Apartheid: Segregation and the Making of the Underclass.* Harvard University Press.

Rieder, Jonathan. 1985. *Canarsie: The Jews and Italians of Brooklyn against Liberalism.* Harvard University Press.

Schelling, Thomas. 1971. "Dynamic Models of Segregation." *Journal of Mathematical Sociology* 1:143-86.

Shevky, Eshref, and Wendell Bell. 1955. *Social Area Analysis: Theory, Illustrative Application and Computational Procedures.* Stanford University Press.

Shlay, Anne, and Peter Rossi. 1981. "Putting Politics into Urban Ecology: Estimating Net Effects of Zoning." *Urban Affairs Annual Reviews* 21:257-86.

Stearns, Linda, and John Logan. 1986. "The Racial Structuring of the Housing Market and Segregation in Suburban Areas." *Social Forces* 65:28-42.

Taeuber, Karl, and Alma Taeuber. 1965. *Negroes in Cities: Residential Segregation and Neighborhood Change.* Aldine.

Van Valey, Thomas L., Wade Clark Roof, and Jerome E. Wilcox. 1977. "Trends in Residential Segregation: 1960-1970." *American Journal of Sociology* 82:826-44.

Villemez, Wayne. 1980. "Race, Class and Neighborhood: Differences in the Residential Return on Individual Resources." *Social Forces* 59:414-30.

White, Michael, Ann Biddlecom, and Shenyang Guo. 1993. "Immigration, Naturalization, and Residential Assimilation among Asian Americans in 1980." *Social Forces* 72:93-118.

Wilson, Kenneth, and Alejandro Portes. 1980. "Immigrant Enclaves: An Analysis of the Labor Market Experiences of Cubans in Miami." *American Journal of Sociology* 86:295-319.

Zhou, Min. 1992. *Chinatown: The Socioeconomic Potential of an Urban Enclave.* Temple University Press.

Attitudes on Residential Integration: Perceived Status Differences, Mere In-Group Preference, or Racial Prejudice?*

LAWRENCE BOBO, *University of California, Los Angeles*
CAMILLE L. ZUBRINSKY, *The Ohio State University*

Abstract

Residential segregation, especially of blacks from whites, remains the common pattern in urban America. This research examines the part that popular attitudes on residential integration may play in the process of residential integration/segregation. Using data from a large multiethnic sample survey in Los Angeles, we examine three hypotheses about the nature of attitudes toward residential integration. The perceived economic status difference hypothesis holds that attitudes about racial residential integration rest upon assumptions about likely class background differences between ethnic groups. The mere in-group preference hypothesis suggests that ethnocentrism results in mutual across-group preferences for residential contact with in-group members. The prejudice hypothesis suggests that hostile attitudes toward an out-group shape views on residential integration. Little evidence in support of the perceived economic status difference and mere in-group preference hypotheses can be found. Theories of prejudice, in particular Blumer's theory of group position, provide much greater leverage on residential integration attitudes. We discuss the implications of the results for actual behavior and aggregate patterns of racial residential segregation.

Racial residential segregation is arguably the "structural linchpin" of American race relations (Bobo 1989; Pettigrew 1979). Analyses of 1980 census data showed that in 16 large metropolitan areas — including Los Angeles — blacks were "hypersegregated" from whites, exhibiting extreme isolation across at least four of five standard indicators (Massey & Denton 1989). Although some modest improvement occurred between 1980 and 1990, blacks remain highly segregated from whites and substantially more isolated from whites than either Asians or Hispanics (Farley & Frey 1994).

* *The authors wish to thank Mia Tuan and two anonymous Social Forces reviewers for their comments on an earlier version of this paper. Please direct correspondence to Lawrence Bobo, Department of Sociology, University of California, Los Angeles, 405 Hilgard Avenue, Los Angeles, California, 90095-1551.*

There are strong reasons to believe that interethnic attitudes play a part in the problem of residential segregation. In particular, the distinctly high rates of black-white segregation would appear to call for attention to potential structures of racial attitude and preference in the population. As Massey and Denton (1989:389) argued: "Blacks are thus unique in experiencing multidimensional hypersegregation. The contrast between them and Hispanics is not easily explained by different socioeconomic characteristics, varying population sizes, different regional locations, or contrasting metropolitan conditions. Although our models cannot eliminate the view that some unmeasured objective factor accounts for the discrepancy between blacks and Hispanics, the models lend credence to the view that blacks remain the object of significantly higher levels of Anglo prejudice than Hispanics." Indeed, the degree of black residential isolation is so unique and persistent that it prompted Massey and Denton (1993) to label the phenomenon "American Apartheid," concluding that racial residential segregation is central to the development and persistence of an urban underclass that is disproportionately African American.

Others have challenged the view that uniquely potent levels of antiblack prejudice and discrimination contribute to the higher levels of black segregation from whites. For example, a number of econometric analyses emphasize such factors as economic status differences, job location, and other differences in tastes as contributing to racial residential segregation (Berry 1979; Leven et al. 1976). A central factor may be mutual patterns of in-group preference as opposed to out-group avoidance or hostility. Clark (1986, 1992) argued that the degree of joint preference among whites and blacks for living near substantial concentrations of coracial/coethnic group members is probably a larger factor in explaining modern residential segregation than white avoidance and discrimination.

Our broad objective in this research is to understand the climate of attitudes, beliefs, and opinions about racial residential segregation. We bring three objectives to the research. First, we wish to characterize the current levels of expressed preference for social distance from members of different racial and ethnic group backgrounds. Such preferences have implications for likely patterns of individual and institutional discrimination in the housing market. Second, we specifically test social psychological aspects of the perceived social class differences, the mere in-group preferences, and the prejudice models as they might influence patterns of residential segregation. Third, in recognition of the increasing diversity of many major metropolitan areas, our research is multiethnic in scope, examining the attitudes of blacks, whites, Hispanics, and Asians.[1] One possibility is that increasing ethnic diversity creates a "buffer," opening to blacks more opportunity for residential mobility and contact with whites.

Views on Racial Residential Segregation

Many factors may contribute to and constrain an individual's choice of housing location: cost and affordability, location, the quality of housing stock, proximity to work, stage in the life cycle, quality of schools, and so on (Galster 1988).

Aggregate patterns of segregation by racial/ethnic group membership can thus stem from any of several individual-level processes. With respect to understanding patterns of specifically racial residential preferences, three hypotheses are typically considered. These hypotheses are, first, that perceived differences in socioeconomic status that heavily coincide with racial/ethnic boundaries contribute to racial residential preferences; second, that members of all social groups tend to be ethnocentric, expressing preference for association and interaction with fellow racial/ethnic group members; and third, that more active out-group avoidance or domination are at the root of racial residential preferences. We elaborate on each of these models below.

THE PERCEIVED SOCIAL CLASS DIFFERENCE HYPOTHESIS

According to this view, racially/ethnically segregated neighborhoods are the result of perceived group differences in socioeconomic status characteristics: income, occupation, and associated differences in life-style (see Jackman & Jackman 1983 on class identities as involving life-style considerations). These perceived status differences are influential both within and across groups such that, "except for the genuinely poor, all people — white and black, rich and not so rich, are willing to pay, and substantially, to avoid class integration. . . . Rising above humble origins to make it in the new and better neighborhood is central to our societal tradition. Without passing judgment on it we must acknowledge the tradition, and we certainly do not seek policies to destroy it" (Leven et al. 1976:202-3). Clark (1988) draws attention to a number of social class characteristics that would increase opposition to having blacks as neighbors. He argues that, in addition to differences in wealth, black households are more likely to have a female head, unemployed adults, and more residents per household. These characteristics leave black home seekers at a disadvantage, not because of their race, but because of class-based differences in the potential for wear and tear to property, lower incomes, and unstable employment patterns.

Thus, under this model, minority group members — particularly blacks — are said to live in segregated areas because they are perceived as lacking the material and cultural-class-based resources needed to obtain housing in more desirable, predominantly white suburban areas. Under this hypothesis, we should find that the more individuals perceive that members of an out-group lag behind his or her own racial/ethnic group socioeconomically, the greater the objection to substantial residential contact with that group.

Despite the commonsense appeal and plausibility of this hypothesis, there is little evidence supporting it. Galster's (1988) comprehensive review concluded that affordability and other race-neutral economic factors (e.g., job location) accounted for at best a small fraction of black-white residential segregation. The bulk of evidence supporting this perspective comes from survey questions in which respondents are asked to explain why groups — usually blacks and whites — tend to live in separate neighborhoods, and/or the extent to which groups are perceived as having the ability to pay for housing in particular areas (Farley et al. 1993; Farley et al. 1978). Black-white comparisons of actual housing expenditures indicate a great deal of overlap, suggesting that many blacks can, indeed, afford to live in "desirable" neighborhoods (Farley et al. 1993).[2]

Likewise, inaccurate knowledge of the cost of housing cannot account for black-white levels of segregation. In 1976, blacks and whites in the Detroit area possessed accurate knowledge of housing prices in the greater Detroit area; by 1992, both blacks and whites tended to overestimate the cost of suburban housing (Farley et al. 1993:11-12). Despite this change, blacks and whites reported extensive and accurate knowledge of blacks' financial capabilities, with both groups recognizing that significant numbers of blacks can, indeed, afford to live in suburban areas (14).

THE MERE IN-GROUP PREFERENCE HYPOTHESIS

Contrary to analyses of residential segregation that give center stage to "white avoidance" or to institutional discrimination (Pearce 1979; Yinger 1986), Clark (1992) asserts the primary importance of in-group preferences. By in-group preferences he refers to "strong desires for own-race combinations in the ethnicity of neighborhoods" (451). These preferences are interpreted as a simple and natural ethnocentrism rather than anti-out-group sentiment or an effort to preserve relative status advantages. Clark argues that there are "strong similarities in own-race preferences among different ethnic groups" (463-64), and that "discrimination in the housing market plays only a minor role" (452). The preferred racial/ethnic composition of a neighborhood is simply one of many characteristics taken into consideration when searching for housing. What is more, under this model, preferences themselves are driven by positive feelings about one's own group, not negative feelings about out-groups. These ethnocentric feelings act in concert with group differences in economic status to explain the bulk of residential segregation: segregation results from the mutually ethnocentric social preferences of consumers (Clark 1986:108-9).[3] Under this hypothesis, strong feelings of in-group attachment or affect should increase objections to substantial residential integration with members of other groups.

Two immediate problems arise with the in-group preference hypothesis. While many individuals may attribute segregation to mutual ethnocentrism, such an account may function as a convenient mask or excuse for those who do not wish to express out-group prejudices more directly. Farley et al. (1994) asked respondents to the 1992 Detroit Area Study to explain why blacks and whites generally lived in different neighborhoods. Variants of the "mere in-group preference" argument were the most common answer, with 30% of whites opining that "birds of a feather flock together" or that "people just prefer it that way." Detailed analysis, however, shows that large fractions of white DAS respondents expressed negative stereotypes of blacks. They were especially inclined to do so when asked why they would not live in neighborhoods with large numbers of blacks.

In addition, emphasizing in-group preferences — positive in-group affect or ethnocentrism — as opposed to prejudice to explain persisting racial residential segregation minimizes the extent to which the preferences of one group constrain the preferences of other groups, particularly blacks. If more than token numbers of black neighbors are "seen as threatening to white households" (Clark 1991:3), and this threat results in avoidance behavior or "white flight," this would seem to indicate the presence of something more closely resembling

negative out-group affect. If preferences for same-race neighborhoods were equal across groups, and were accompanied by assurances of equal quality of life across groups, it might be persuasive to view in-group preferences as neutral. However, because of the manifestly racist historical origins of residential segregation in the U.S. (Cell 1981; Jaynes & Williams 1989; Woodward 1974) and because its effects are not neutral in their consequences (Massey & Denton 1993), it seems shortsighted to consider them neutral.

THE PREJUDICE HYPOTHESIS

Two variants of the prejudice hypothesis are relevant to attitudes toward residential integration. The first is closer to *traditional prejudice* and, in contrast to the mere in-group preference hypothesis, stresses the importance of out-group hostility in determining individual attitudes about residential contact (Allport 1954; Katz 1991; Pettigrew 1982). Prejudice is typically defined as an irrational antipathy against minority groups and their members. Prejudice is understood as heavily imbued with negative affect and negative stereotypes that make the views of the prejudiced individual unreceptive to reason and new information (Jackman 1994). Accordingly, under the traditional prejudice hypothesis we should find that measures of negative affective response to out-group members and of negative stereotypes should be the most strongly associated with objections to residential integration.

A second variant of the prejudice hypothesis is rooted in Blumer's (1958) theory of race prejudice as *a sense of group position*. Rather than placing negative feelings and beliefs at the core of prejudice, Blumer argued that prejudice involves a commitment to a specific group status or *relative* group position. The group position hypothesis suggests that neither mere in-group preference nor out-group hostility are sufficient to give prejudice social force. Instead, what matters is the magnitude or degree of difference that in-group members have socially learned to expect and maintain relative to members of specific out-groups. As a result, under the sense of group position hypothesis, it is the degree of difference between in-group attachment and out-group hostility that should be most strongly associated with objections to residential integration (see Bobo 1988). The greater the affective differentiation from members of an out-group, with such differentiation understood as one indicator of a preferred superior group position, the greater the likelihood of objections to residential integration.

Both the traditional prejudice hypothesis and group position as prejudice hypothesis view residential segregation as connected to attitudes about an out-group. As such, racial residential segregation persists because "whites prefer and are willing to pay more for segregation than blacks are willing [or able] to pay for integration" (Muth 1986:9). Several types of research on attitudinal data point to the plausibility of the prejudice hypothesis. To be sure, there has been a sweeping rise in whites' advocacy of the *principle* of free residential choice for blacks. Trend studies have shown, however, that whites still express great opposition to the enforcement of blacks' right to live wherever they can afford (Bobo, Schuman & Steeh 1986; Pettigrew 1973, 1979; Schuman, Steeh & Bobo 1985). Experimental data have also shown many whites to express greater

residential social distance from blacks than from Asians (Schuman & Bobo 1988).

Data and Measures

The data for testing these several views of residential integration attitudes come from the 1992 Los Angeles County Social Survey (Institute for Social Science Research 1992; hereafter LACSS). The LACSS is a countywide telephone survey of adults living in households selected by random dialing of digits. In addition to the general countywide sample, oversamples of telephone numbers in ZIP code areas of high black concentration (65% or more) and of high Asian concentration (30% or more) were used to efficiently generate larger numbers of black and Asian respondents. To fully capture the views and opinions of Los Angeles's very large Latino population, a Spanish language translation of the questionnaire was developed. Monolingual Spanish speakers and those preferring to conduct the interview in Spanish were interviewed in Spanish.

There were a total of 1,869 respondents, with 625 white, 483 black, 477 Latino, and 284 Asian respondents. The split-ballot design for some measures resulted in fewer cases for portions of the analysis.

Interviews were conducted by trained student interviewers taking part in a survey research methods course and by the regular interviewing staff of the Survey Research Center. The student interviewers received 12 hours of training. The LACSS procedure is to attempt a telephone number 12 times, systematically varying the day of the week and time of day before dropping it from the sample. The questionnaire averaged 38 minutes in length. The study had an overall cooperation rate of 55%. Within each major racial/ethnic group the distribution of sample characteristics on key social background factors closely resemble data from the 1990 census (see Bobo et al. 1992).[4] A systematic analysis of potential nonresponse bias, based on the procedure recommended by O'Neil (1979), indicated no pattern of significant nonresponse bias (Greenwell, Strohm & Bobo 1994).

Results

RESIDENTIAL INTEGRATION ATTITUDES

In order to tap views on racial residential integration, the 1992 LACSS included a series of questions that asked each respondent whether he or she would strongly favor, favor, neither favor nor oppose, oppose, or strongly oppose living in a neighborhood where half of their neighbors would be (in the case of white respondents) blacks, or Latinos, or Asians. Members of each major racial/ethnic category were asked about each potential out-group. In the current research, we choose to ask about a neighborhood composed of "half" members of another racial/ethnic group in order to effectively pose the issue of the importance of race/ethnic makeup to the respondent. Mentioning a half other race/ethnicity neighborhood immediately suggests substantial integration but does not go the further step of putting the respondent in a minority status,

which would raise an additional conceptual issue. The distribution of responses, combining "oppose" and "strongly oppose" responses, are shown by race/ethnicity of respondent and by target group in Table 1.

We should first note that, in an absolute sense, at least compared to actual levels of racial residential segregation, these numbers suggest substantial openness to residential integration in the mass public. For example, as Table 1 shows, just 34% of whites said that they were opposed or strongly opposed to living in a neighborhood where *half* of their neighbors would be black. Another 46% offered neutral opinions, and 1 in 5 said they would favor or strongly favor living in such a highly integrated area. To a degree, this may reflect a bias of socially desirable responses, with many respondents, whites in particular, unwilling to voice objections to residential contact with minorities. Yet we do not believe the pattern reflects a serious response bias of this kind, nor does it constitute a sharp disjuncture with actual patterns of racial residential segregation. Respondents did not give unreflective pro-integration responses. First, and most important, it is immediately apparent that members of each group react differently to each minority group. The rate of whites' objections to residential integration with Asians, for instance, is 12% lower than that observed when asked about blacks. Similarly, the rate of Asian objections to residential contact with blacks is nearly twice as high as that for potential contact with Hispanics and about six times as high as that for potential contact with whites. Second, awareness of the general perception of blacks as the least desirable neighbors, a perception reflected in these results, may be more consequential for individual housing choices and location than a deeply held personal aversion. Blacks may be discouraged by the unpredictable reaction from a relatively small number of not easily identified, potentially hostile whites from moving into predominantly white neighborhoods. At minimum, in rank order of groups, the attitudinal data closely mirror the findings from more objective measures.

Several other patterns emerge. Only a trivial percentage of blacks, Hispanics, and Asians express objection to living in a largely white neighborhood. The figure is below 10% for each of the minority groups. This is consistent with the assumption that greater residential contact with whites is typically seen as being associated with upward social mobility and access to better neighborhoods and services.

Blacks are clearly the least preferred potential neighbors. The highest rates of objection to interracial residential settings occurs when blacks are the target group of potential contact. Fully 46% of Asians object to living in a half black neighborhood and a least a third of both whites and Hispanics express similar objections. If viewed as a rank order of preferences in the population as a whole, then, blacks are unequivocally at the bottom of the preference hierarchy and whites just as unequivocally are at the top.

Black respondents are the least likely to object to residential integration. White respondents are the most likely to object to interracial residential contact, with but one exception (i.e., Asians are more strongly opposed to contact with blacks). All in all, these patterns strongly suggest the presence of a well-defined racial rank order with respect to housing. In this racial order, blacks are the bottom group, at once most open to integrated living with members of other groups and yet confronting the most resistance from other groups.

TABLE 1: Race and Residential Integration Attitudes Objections to Residential Integration with Out-Groups, by Race/Ethnicity

	Respondent Race			
	Whites	Blacks	Hispanics	Asians
Target group				
Whites	—	8.3%	8.6%	6.1%
	—	(481)	(476)	(281)
$\chi^2 = 16.62$				
df = 8				
$p < .05$				
Blacks	34.3	—	32.8	46.2
	(613)	—	(473)	(277)
$\chi^2 = 32.83$				
df = 8				
$p < .0001$				
Hispanics	25.6	21.4	—	24.9
	(619)	(221)	—	(278)
$\chi^2 = 69.63$				
df = 8				
$p < .0001$				
Asians	22.7	19.2	20.6	—
	(618)	(478)	(472)	—
$\chi^2 = 47.00$				
df = 8				
$p < .0001$				

In order to understand the social bases of these attitudes, we examined the percentage objecting to residential integration by race of respondent, race of target group of the question, and several major social background variables (respondent age, education, occupation, sex, family income, and nativity). In general, these results suggest infrequent and small effects of these background factors on residential integration attitudes. Out of 72 tests, only 21 attain conventional criteria of significance.[5]

Nonetheless, several of these patterns are worthy of note. First, among white respondents, the data consistently show that the lower the level of education, the higher the likelihood of objecting to residential integration. Second, there is a slight tendency for younger white and black respondents to express greater opposition to contact with Hispanics and Asians than do their older coethnics. Third, nativity sometimes influences the residential segregation views of Hispanics and Asians, though in opposite directions. Foreign-born

Asians are highly likely to object to residential integration with blacks. Native-born Latinos are more likely to object to residential integration with blacks and whites than are foreign-born Latinos.

IN-GROUP PREFERENCE HYPOTHESIS

Our primary analytical task in the remainder of this article is to understand the possible sources of individuals' racial residential integration attitudes. Clark (1986, 1992) argued that in-group preference was a substantial factor in racial residential preferences. We measured in-group attachment with a widely used procedure known as a feeling thermometer. Respondents were asked to rate each of several social groups on a scale running from 0 degrees, which represented extremely cold feelings, to 100 degrees, which represented extremely warm feelings, and where 50 represented neutral feelings. This approach to tapping affective reactions to issues, candidates, and social groups has been widely used in the literature on intergroup attitudes within sociology (Bobo 1988; Jackman 1977; Jackman & Muha 1984; Schuman, Steeh & Bobo 1985), and political psychology (Sears 1988). We asked respondents to rate their own group, and each of the other major racial/ethnic groups on the thermometer scales. The mean scores by race are shown in Table 2.

Consistent with Clark's "mere in-group preference" hypothesis, all groups express in-group preferences. That is, the in-group thermometer rating is on average the highest absolute rating given by members of each group (highest scores run along the diagonal). Furthermore, by a slight margin blacks have the highest in-group affective rating — 71.2 — followed by Hispanics at 68.3, then whites at 66.3 and Asians at 65.0. A pattern worthy of note is that minorities consistently give whites higher affective ratings than they do to other out-groups. In other words, the highest average out-group rating among minorities comes in reaction to whites, not in reaction to members of another minority group. This pattern again points to the presence of an American racial rank order, with whites *consensually* regarded as occupying the most preferred social position.

These thermometer measures allow us to test aspects of the mere in-group preference hypothesis. Table 3 shows the Pearson correlation coefficient for the in-group thermometer rating and each of the racial residential integration questions by respondent race/ethnicity. As is immediately evident, in-group affect has a uniformly weak relation to racial residential integration attitudes. In only one instance does the correlation rise above .10 (blacks' reactions to Hispanics, $r = -.13$), and in this case the effect is in the wrong direction, suggesting that as positive in-group feeling increases among blacks, opposition to residential integration with Hispanics tends to decrease. This first-cut analysis suggests that mere in-group preference is not likely to be a substantial determinant of whites' reactions to racial residential integration or to the reactions of blacks, Hispanics, or Asians either.

TABLE 2: Race and Mean Feeling Thermometer Ratings Thermometer Scales[a]

	Respondent Race				
	White	Black	Hispanic	Asian	f
Group rated					
Whites	66.3	61.4	62.8	62.1	6.03***
Blacks	58.4	71.2	55.4	53.1	59.50***
Hispanics	58.8	58.9	68.3	53.7	32.50***
Asians	58.8	55.2	55.5	65.0	14.96***

[a] Scores are means on a 0 to 100 scale where high scores indicate warmer or more positive feelings.

*** $p < .001$

IS IT PREJUDICE?

The thermometer ratings also allow us to test aspects of the two prejudice-based hypotheses about attitudes on racial residential segregation. It is possible that affective hostility to an out-group is a more substantial element of racial residential preferences than mere in-group preference. To examine this possibility, the second row of figures in each target-group block of Table 3 reports the correlation between racial residential integration attitudes and the appropriate out-group thermometer rating. Among all groups, the out-group affective rating is more strongly correlated with residential integration attitudes than mere in-group preference. The general pattern is that as the out-group affective rating improves, opposition to interracial residential contact with members of that group tends to decline.

We can also test aspects of the group position model of prejudice as a basis of racial residential integration attitudes. Under this model, it is ideas and beliefs about the relative group positions that constitute the psychological core of prejudice, not merely in-group attachment or out-group aversion. We attempt to capture such positional commitments with a difference score that involves subtracting an individual respondent's rating of each out-group from his or her in-group rating using the thermometers. We refer to this measure as a difference score or affective differentiation measure. The third row of each target-group block of figures shows the correlation between the affective differentiation measure and the racial residential integration attitudes measure.

The affective difference measure is the strongest correlate of racial residential integration attitudes among whites. In each case, for white respondents the correlation exceeds .3, and it reaches .43 in the case of reactions to living in a half Asian neighborhood. In short, as the affective difference that whites prefer to maintain between themselves and members of minority groups rises, so does the level of opposition to racial residential integration. Consistent with the group position model then, there appears to be something status-oriented

TABLE 3: Correlation of In-Group Affect, Out-Group Affect, and Affective Differentiation with Residential Integration Attitudes[a]

Target group		Respondent Race: Whites	Blacks	Hispanics	Asians
Whites	In-group	—	-.03	-.04	-.01
	Out-group	—	-.25***	-.16***	-.19**
	Difference	—	.18***	.11*	.18**
Blacks	In-group	.09*	—	-.08	-.02
	Out-group	-.25***	—	-.17***	-.28***
	Difference	.37***	—	.09	.25***
Hispanics	In-group	.04	-.13**	—	-.07
	Out-group	-.26***	-.30***	—	-.27***
	Difference	.32***	.17***	—	.20***
Asians	In-group	.10*	-.04	-.05	—
	Out-group	-.32***	-.36***	-.18***	—
	Difference	.43***	.30***	.11*	—

[a] Figures are Pearson's correlation coefficients.

* $p < .05$ ** $p < .01$ *** $p < .001$

and positional underlying whites' reactions to the possibility of substantial racial residential integration, not simply in-group preference or out-group aversion.

The affective differentiation measures, although consistently stronger correlates of residential integration attitudes than mere in-group preference, usually exhibit weaker correlations than simple out-group ratings among minority respondents. For blacks, Hispanics, and Asians the correlation between the affective differentiation measure and residential integration attitudes is always weaker, and sometimes substantially so (e.g., .17 versus -.30 for blacks' reactions to Hispanics), than the correlation with the out-group affective rating alone. This immediately implies that status and positional issues have less to do with how minority respondents form their attitudes on residential integration than is apparently true of whites.

We wished to gain greater leverage on racial prejudice as a possible source of attitudes on residential integration. To do so, we also examined how measures of racial stereotypes relate to the racial residential integration attitudes.[6] Table 4 shows, by race/ethnicity of respondent and of the target group of the question, mean scores on a series of bipolar trait rating scales. Each item and the overall scales range from 1 to 7, with a score of 4 as a neutral or

mid point. Each item has been scored so that higher scores reflect more negative ratings on the dimensions of intelligence, welfare dependency, and difficulty to get along with socially. Time constraints made it impossible to ask about a much wider set of traits, but these three were chosen on substantive grounds. Intelligence is a classic issue in intergroup relations. In an achievement-oriented society such as the U.S., intelligence is often credited with explaining socio-economic success or failure and a host of other social and behavioral outcomes, such as involvement in crime. Welfare dependency has been a longstanding aspect of antiblack stereotypes in the U.S. and may be increasing as an aspect of stereotypes of Latinos as well. The "hard to get along with" dimension was included because of the widespread discussion that many recent Asian immigrant groups, especially but not exclusively Koreans, have brought cultural styles of interaction that are more brusque than is typical of U.S. culture. Hence, we expect that this set of traits taps critical dimensions of minority group stereotypes that are likely to have bearing on willingness to share residential space with minority group members.

The rows of Table 4 within each target-group block report the mean rating on each of three trait dimensions, and the fourth row reports an average stereotype score based on all three traits. The figures in parentheses report the Pearson correlation for each trait measure and the stereotype indexes with the racial residential integration attitudes. Among whites in particular, the results suggest that negative stereotypes are an element of how individuals form their attitudes to racial residential integration.

Consider first just the stereotypes themselves. Blacks receive the most negative overall ratings and whites, predictably, receive the most favorable overall ratings. Asians and Hispanics tend to fall in between these extremes, with ratings of Hispanics quite close to those found for blacks and ratings of Asians quite close to those found for whites. Among white, Asian, and Hispanic respondents, blacks receive the highest average negative stereotype ratings. The single most negative rating of blacks among all groups occurs on the welfare-dependency trait. There is also a tendency for welfare dependency to be the negative trait most attributed to Hispanics as well. The most negative rating of Asians occurs consistently on the "hard to get along with" dimension, as expected.

Stereotypes are usually correlated with racial residential integration attitudes. As stereotypes become more negative, opposition to residential integration tends to increase. This pattern is strongest and most consistent among white respondents, but there is also a trend in this direction among blacks, Hispanics, and Asians.

IS IT PERCEIVED SOCIAL CLASS DIFFERENCES?

Before turning to multivariate analyses, we consider the bivariate relation of perceived socioeconomic status differences and attitudes on residential integration. We tapped perceived socioeconomic status differences with two measures. The first is a simple bipolar trait-rating question that asked respondents to rank each major racial/ethnic group on a 1 to 7 continuum where 1 meant most group members "tended to be rich" and 7 meant most group

TABLE 4: Race and Mean Stereotype Trait Ratings[a]

	Respondent Race			
Target group	Whites	Blacks	Hispanics	Asians
Whites				
Unintelligent	3.31	3.18 (.11)	2.84 (.18)	3.07 (.11)
Prefer welfare	2.67	2.97 (-.02)	2.96 (.01)	2.72 (-.00)
Hard to get along with	3.08	3.50 (.18)	3.10 (.18)	3.07 (.07)
Stereotype rating	3.02	3.21 (.12)	2.97 (.17)	2.95 (.08)
Blacks				
Unintelligent	4.05 (.20)	3.60	3.77 (.10)	4.27 (.28)
Prefer welfare	4.10 (.29)	3.98	5.22 (.10)	4.84 (.28)
Hard to get along with	3.81 (.19)	3.57	4.55 (.18)	4.02 (.25)
Stereotype rating	4.00 (.31)	3.72	4.51 (.19)	4.37 (.35)
Hispanics				
Unintelligent	3.98 (.19)	4.05 (.09)	3.50	4.07 (.20)
Prefer welfare	3.86 (.19)	4.30 (.07)	4.00	4.47 (.11)
Hard to get along with	3.52 (.26)	3.69 (.09)	2.98	3.66 (.16)
Stereotype rating	3.79 (.29)	4.02 (.11)	3.50	4.06 (.21)
Asians				
Unintelligent	3.29 (.07)	3.30 (-.06)	2.88 (.03)	3.04
Prefer welfare	2.74 (.20)	3.01 (-.03)	3.21 (.08)	2.89
Hard to get along with	3.82 (.25)	4.24 (.24)	4.07 (.11)	3.18
Stereotype rating	3.29 (.25)	3.52 (.08)	3.39 (.11)	3.03

[a] Figures are means on a 1 to 7 scale where 7 is the negative end of a bipolar rating continuum. Figures in parentheses are zero-order correlations of the stereotype measure with residential integration attitudes.

members "tended to be poor." The second measure is a difference score between a respondent's in-group rating on this measure and his or her rating of each out-group. A score of zero on this measure would thus indicate a perception of no average economic status difference between groups, positive scores would indicate a perception that one's in-group was of higher economic status than the out-group, and negative scores on the difference measure would indicate a perception that one's in-group members were of lower economic status. Results for both measures are shown in panel A of Table 5. Panel B reports the correlation of each measure, by respondent race and target group race/ethnicity, with racial residential integration attitudes.

All respondents tend to perceive a similar pattern of economic status differences between groups. In broad terms, these perceived differences conform to actual average patterns of difference in economic status (Hirschman 1983).

TABLE 5: Perceived Status Differences and Residential Integration Attitudes

Panel A: Mean Group Rating[a]

	Respondent Race			
	Whites	Blacks	Hispanics	Asians
Target group				
Whites				
Mean group rating	3.50	3.10	3.03	3.00
In-group/out-group difference	—	-1.95	-1.46	-.43
Blacks				
Mean group rating	5.04	5.03	4.71	5.17
In-group/out-group difference	1.55	—	.22	1.74
Hispanics				
Mean group rating	5.00	5.20	4.49	4.82
In-group/out-group difference	1.51	.17	—	1.40
Asians				
Mean group rating	3.51	3.65	3.33	3.42
In-group/out-group difference	.01	-1.38	-1.16	—

[a] Figures in the upper row are means on a 1 to 7 scale where 1 means "rich" and 7 means "poor." Figures in the lower row are differences scores between in-group rating and out-group rating.

Thus, blacks and Hispanics are the groups most likely to be rated toward the poor end of the continuum and Asians and whites most likely to be rated toward the rich end of the continuum. The perceived in-group versus out-group difference mirrors this pattern as well. For example, whites and Asians both see blacks and Hispanics as lagging behind them economically, but on average see little difference in economic status between one another. As panel B shows, however, neither the absolute perceived economic status nor the perceived status difference is substantially correlated with a respondent's racial residential integration attitudes. Out of 24 possible instances only 4 correlations attain significance at the .01 level, and one of these is in the wrong direction. *In sum, concern about perceived differences in economic status is unlikely to be a powerful component of how respondents form their attitudes on racial residential integration.*

It is important to determine whether the bivariate patterns summarized to this point persist in a fuller multivariate analysis. There are important differences among respondents and across groups in educational attainment, income, sex distribution, and nativity. Therefore we conducted a series of OLS regression analyses that controlled for these factors and then introduced, in separate

TABLE 5: Perceived Status Differences and Residential Integration Attitudes

	Panel B: Correlations[b]			
	Respondent Race			
Target group	Whites	Blacks	Hispanics	Asians
Whites				
Group rating	—	-.01	.11*	.06
In-group/out-group difference	—	.01	.02	.01
Blacks				
Group rating	.02	—	.03*	.18
In-group/out-group difference	.02	—	.01	.13
Hispanics				
Group rating	.07	.11*	—	.09
In-group/out-group difference	.03	.06	—	.03
Asians				
Group rating	.03	-.14*	.03	—
In-group/out-group difference	.02	-.07	.02	—

[b] Figures are Pearson correlations with the residential integration attitudes measure.

* $p < .05$ ** $p < .01$ *** $p < .001$

models, the in-group affect measure, the out-group affect measure, and the affective differentiation measure. Each model included the perceived economic status difference measure and the stereotype index rating. These results are shown in Table 6.

In only 3 of 12 possible instances did the in-group preference measure exhibit a significant relationship with the residential social distance measure. Of note, two of these occur for whites and involve reactions to contact with the *racially* distinct out-groups: blacks and Asians. This pattern confirms the longstanding sociological observation that differences perceived as racial rather than ethnic in nature are often more socially divisive (Lieberson 1980; Stone 1985). Simple in-group favoritism is, plainly, not a powerful determinant of social distance attitudes. It has no effect whatever among Hispanic or Asian respondents and works in the wrong direction (according to the preference model) for blacks' reactions to Hispanics.

The perceived SES gap measures, with only one exception, are not related to social distance feelings. The one instance of a significant effect, ironically, is in the direction opposite to expectation. The larger the gap blacks see between

TABLE 6: Multivariate Regression Coefficients Examining the Effects of Perceived SES Gap, Affective Ratings, and Stereotypes on Residential Integration Attitudes

	In-Group Affect	Out-Group Affect	Difference	Perceived SES Gap	Stereotype Rating	R^2
			White Respondents			
Target group						
Blacks	.11***	—	—	-.00	.29***	.11
	—	-.17***	—	-.01	.27***	.13
	—	—	.31***	-.01	.25***	.19
Hispanics	.06	—	—	.01	.26***	.07
	—	-.22***	—	.02	.22**	.11
	—	—	.29***	.01	.20**	.15
Asians	.12**	—	—	.00	.24***	.09
	—	-.27***	—	.02	.17***	.15
	—	—	.38***	.03	.16***	.21
			Black Respondents			
Whites	-.07	—	—	-.06	.17***	.04
	—	-.18***	—	-.04	.13*	.07
	—	—	.08	-.04	.15**	.04
Hispanics	-.12*	—	—	.05	.11*	.06
	—	-.28***	—	.05	.08	.12
	—	—	.15**	.03	.12*	.06
Asians	-.03	—	—	-.12*	-.11*	.04
	—	-.32***	—	-.08	.05	.14
	—	—	.26***	-.08	.08	.11
			Hispanic Respondents			
Whites	-.03	—	—	.00	.15**	.06
	—	-.13**	—	-.00	.14**	.08
	—	—	.09	.00	.15**	.07
Blacks	-.03	—	—	.04	.16**	.03
	—	-.15**	—	.03	.13*	.05
	—	—	.11*	.04	.16**	.04
Asians	-.01	—	—	.01	.09	.03
	—	-.13**	—	.02	.06	.04
	—	—	.10*	.03	.08	.04
			Asian Respondents			
Whites	.00	—	—	-.04	.07	.03
	—	-.15*	—	-.03	.03	.05
	—	—	.15*	-.02	.06	.05
Blacks	-.07	—	—	.08	.31***	.13
	—	-.21***	—	.10	.25***	.17
	—	—	.14*	.09	.26***	.15
Hispanics	-.08	—	—	.08	.18**	.04
	—	-.24***	—	.09	.10	.09
	—	—	.15*	.08	.13	.06

* p < .05 ** p < .01 *** p < .001

themselves and Asians, the more open they are to residential contact with Asians. In short, what an individual assumes about the average economic status of out-groups has little to do with willingness to enter a substantially integrated community.

The various prejudice measures have far more consistent and stronger effects on social distance feelings. The effects of prejudice vary with the measure and, to a degree, with the race of the respondent and the race of the target group.

The most consistent effects are found for the out-group affect measure. Across each of the 12 tests, we found that as affect toward an out-group improved, opposition to residential contact significantly declined. Indeed, even for the two occasions when in-group affect has the anticipated effect, it is clear that the increment to variance explained above the baseline model is greater for the out-group affect measure. Out-group affect matters irrespective of race of the respondent and irrespective of race of the target group.

The affective differentiation measure influences social distance feelings, particularly among white respondents. Although sometimes a significant predictor among black, Hispanic, and Asian respondents, it adds no more to the amount of variance explained and typically less than the out-group affect measure. That is, affective differentiation is a uniquely potent predictor of social distance feelings among white respondents. The implication is that they are maintaining a social status difference, as opposed to expressing mere in-group preference or out-group hostility.

Finally, negative stereotypes are typically significant predictors of social distance preferences. Negative stereotypes are most consistently important among white respondents. The effects are less consistent among minority respondents, especially among Asians, and to a degree, among Hispanics.[7]

ON LINKING ATTITUDES AND BEHAVIOR

We do not have measures of a respondent's recent residential moves or future intentions. Hence, we cannot examine whether change occurred (or might occur) in an attitude-(in)consistent manner. However, the 1992 LACSS did ask respondents to report on the main ethnic group in their neighborhoods, on whether the neighborhood was undergoing ethnic change, and if so, what the nature of the change was. A brief consideration of these data sheds a small amount of light on possible attitude-behavior connections.

Our data on reported neighborhood composition and change are both reassuringly consistent with familiar demographic analysis results and, more important, with the attitudinal patterns reported above. The upper panel of Table 7 shows that blacks are the group most likely to live among coethnics (64.7%) and the least likely to live in largely white areas (4.5%). Likewise, whites are the second most likely to report sharing residential space with coethnics (58.2%) and the least likely to report living in largely black areas (4.9%). Asians are the minority group most likely to report living in largely white areas (20.8%), a pattern supported by the high proportion of whites who report living in largely Asian areas (17.3%). Both Asians and blacks show high

proportions reporting that they live in largely Hispanic areas, 19.4% and 17.7%, respectively.[8]

The second panel shows that blacks are the group most likely to report that their neighborhoods are undergoing ethnic change (61.8%), followed by whites at 56.1%. In general, the experience of ethnic residential change is common in Los Angeles County, with roughly half of Asian and Latino respondents also reporting that their neighborhoods are undergoing change.

The nature of these changes is made a bit clearer in the lowest panel of Table 7. Among those reporting that a change was taking place, blacks overwhelmingly report (77.8%) that their neighborhoods are becoming more Hispanic. The modal response among both Latinos and Asians, however, is that their neighborhoods are experiencing an influx of coethnics, 45.5% and 66.2%, respectively. These patterns presumably reflect the high rates of new immigration for these groups (Waldinger 1989). Whites undergoing change most often report that their neighborhoods are becoming more Asian (50.3%), followed by an increasing number of Hispanics (31.1%). Few whites see their neighborhoods becoming more black. This is a telling contrast with whites' high and increasing frequency of residential contact with Asians. Blacks and Asians now constitute roughly equal proportions of the population in Los Angeles county, but whites — in a pattern directly consistent with the attitudinal results we have reported — are now far more likely to share residential space with Asians than with blacks and are far more likely than blacks to anticipate the substantial future movement of Asians into their neighborhoods.

Although we cannot test the hypothesis that attitudes directly translate into residential choices and aggregate housing patterns, the results in Table 8, which show the association between reported neighborhood composition and our residential integration attitude measure, are suggestive. First and foremost, a significant association between attitude and reported residential neighborhood composition occurs most consistently among white respondents. White respondents living in mainly white (37.2%) or mainly Asian neighborhoods (43.7%) are highly likely to object to residential integration with blacks. Whites from the same two types of communities are also the most likely to object to residential integration with Hispanics. These patterns are consistent with national survey results reported by Sigelman and Welch (1993), who find more positive racial attitudes among whites who have black neighbors.

We are much less likely to observe an association between attitude and reported residential composition among blacks, Asians, and Hispanics than among whites. The only exceptions to this case occur when blacks are the target of the residential integration question among those Latino and Asian respondents who live in coethnic-dominated neighborhoods or in largely white neighborhoods. In particular, Asian respondents living in mainly white or mainly Asian neighborhoods express the absolute highest levels of objection to residential integration with blacks, 52.7% and 50.8%, respectively.

It is impossible to judge from these data whether attitudes led to the current residential locations, or whether residential location comes to affect attitude. Cross-sectional data face a severe constraint in this regard, as others have observed (Jackman & Crane 1986). In all likelihood, there are reciprocal effects

TABLE 7: Respondent Race and Actual Residential Patterns

	Race of Respondent			
	White	Hispanic	Black	Asian
Neighborhood composition				
Mainly white	58.2	12.2	4.5	20.8
Mainly black	4.9	23.9	64.7	6.5
Mainly Hispanic	13.2	40.1	17.7	19.4
Mainly Asian	17.3	15.3	1.3	45.9
Mixed	6.4	8.4	11.9	7.5
Total (percent)	100	99.9	100.1	100.1
	(622)	(476)	(470)	(279)
$\chi^2 = 1069.69$				
df = 12				
$p < .001$				
Neighborhood ethnic composition changing				
Yes	56.1	48.4	61.8	48.9
No	43.9	51.6	38.2	51.1
Total (percent)	100	100	100	100
	(624)	(477)	(482)	(284)
$\chi^2 = 21.78$				
df = 3				
$p < .001$				
Ethnic change				
More black	6.6	11.3	5.7	11.5
More Hispanic	31.1	45.5	77.8	14.4
More Asian	50.3	36.8	5.1	66.2
More white	4.0	4.8	7.7	5.0
More mixed	8.0	1.7	3.7	2.9
Total (percent)	100	100	100	100
	(350)	(231)	(297)	(139)
$\chi^2 = 311.13$				
df = 12				
$p < .001$				

TABLE 8: Percentage Objecting to Residential Integration by Main Racial/Ethnic Group in Respondent Neighborhood

	Respondent Neighborhood Composition			
	Mainly White	Mainly Black	Mainly Hispanic	Mainly Asian
White respondents				
Target group				
Black***	37.2 (352)	6.7 (30)	35.8 (81)	43.7 (103)
Hispanic***	28.8 (351)	10.0 (30)	22.2 (81)	26.4 (106)
Asian**	25.6 (352)	6.7 (30)	24.7 (81)	18.9 (106)
Black respondents				
Target group				
White	0.0 (9)	8.9 (303)	8.5 (82)	16.7 (6)
Hispanic	33.3 (21)	20.1 (303)	21.9 (82)	16.7 (6)
Asian	23.8 (21)	18.9 (301)	19.3 (83)	16.7 (6)
Hispanic respondents				
Target group				
White	5.2 (58)	13.1 (114)	7.4 (190)	4.1 (73)
Black*	33.9 (56)	28.1 (114)	37.4 (190)	27.4 (73)
Asian	26.3 (57)	19.5 (113)	21.3 (188)	15.1 (73)
Asian respondents				
Target group				
White	5.2 (58)	11.1 (18)	3.8 (53)	7.9 (127)
Black*	52.7 (55)	22.2 (18)	45.3 (53)	50.8 (126)
Hispanic	31.0 (58)	22.2 (18)	26.9 (52)	28.0 (125)

* $p < .05$ ** $p < .01$ *** $p < .001$

with attitude shaping location decisions to as great an extent as resources, information, and other considerations allow, while the experience of particular neighborhoods, group relations, and status and service considerations then come to reshape attitudes (Galster 1989; Sigelman & Welch 1993).

The implications of the Table 7 and Table 8 results taken as a whole, however, tend to point in the direction of racial attitudes contributing to housing location decisions and patterns. At minimum, the data suggest that many whites living in white neighborhoods, and other minorities living in white neighborhoods or in coethnic-dominated neighborhoods are likely to resist the

entry of blacks into their communities. These results strongly parallel recent analysis by Massey and Gross (1991). They argued that, despite evidence of declining residential segregation between 1970 and 1980, there still is a very low upper bound on white willingness to share residential space with blacks. Aggregate segregation declines only to the extent that it does not increase the likelihood of black-white contact much above the 1980 average proportion of .053. Likewise, Farley and Frey (1994:40) found that the largest declines in segregation took place "in metropolitan areas in which blacks made up a small percentage of the neighborhood of the typical white."

The extraordinarily high level of black-white isolation is thus unlikely to break down quickly, even while minority groups of comparable proportionate representation in the population make substantial entry, at least relative to blacks, into white communities. The pattern of attitude distribution and dynamics we have documented seems very much a constitutive element of this process of residential mobility for some and continued apartheidlike conditions for others.

Discussion and Conclusions

We began this research with the goal of examining the mere in-group preference, perceived economic status differences, and prejudice models as hypotheses about the nature of racial residential integration attitudes. Our results provide virtually no support for the in-group preference or the perceived economic status difference hypotheses. We do find general patterns of ethnocentrism, yet ethnocentrism alone is a small component of attitudes on residential integration, irrespective of the race of the target group or of the respondent. We do find consensually held perceptions of racial/ethnic group differences in average economic status, yet these perceived differences in status play almost no role in residential integration attitudes. In short, in order to understand attitudes on racial residential integration we must be expressly concerned with views about other racial/ethnic groups.

Prejudice does appear to be an important element of how respondents form their views on residential integration. The type or form of prejudice that shapes residential integration attitudes appears to depend, however, on majority or dominant group status versus minority or subordinate group status. We find that among white respondents the affective differentiation measure is a strong predictor of residential integration attitudes. Its effects always exceed those of in-group preference and of out-group hostility. In addition, the stereotyping measure is consistently influential among white respondents, irrespective of the minority target group. Among minority respondents it is more often the case that out-group hostility exceeds affective differentiation in importance. What is more, the effects of stereotyping are a good deal less consistent for minority respondents.

We believe that the best substantive reconciliation of these incommensurate patterns is to acknowledge that the U.S. has a relatively clear-cut racial/ethnic hierarchy, or racial order, at least at its top and bottom ranks. The historical and present dominant social group is white Americans. The historical (and at least

perceptually, if not also in fact) present bottom group is African Americans. (See Jaynes & Williams 1989; Massey & Denton 1993.) For blacks, Latinos, and Asians, economic and social advancement is associated with greater proximity and similarity to white Americans. For whites, integration — especially with blacks — brings the threat of a loss of relative status advantages. As a result, attitudes on an issue like racial residential integration are likely to have very different meanings to whites than to members of any of the minority groups, even comparatively affluent Asians. Until such time as Asians have long occupied a status of comparable socioeconomic status with whites, even members of this group should view the prospect of residential contact with whites as upward mobility. Whites are the group, then, most likely to view any increased racial residential integration — with any other group — as changing traditional status relations of relative dominance and privilege. That is, they tend to view integration with any of the minority groups as threatening or undermining a previous status relation of superiority. It is for this reason, we speculate, that the affective differentiation measure assumes greater importance to the attitudes of whites than it does for any of the minorities. Minorities cannot view one another as threats to a traditionally privileged status in quite this fashion. Instead, if minorities express aversion to residential contact with members of other groups, it tends to be based more directly on affective hostility to that group, not an effort to preserve a longstanding relation of privileged status.

Our analysis can be read both positively and negatively. Two aspects of our results suggest that racial residential segregation may continue its gradual decline. First, average levels of openness to substantial integration, even among whites in reaction to large numbers of black neighbors, was fairly high. Second, the basic human tendency toward ethnocentrism, although evident in our data, has comparatively little to do with attitudes on racial residential integration. Both these patterns are consistent with the general trend toward more tolerant racial attitudes (Schuman, Steeh & Bobo 1985) and modest declines in actual levels of residential segregation (Farley & Frey 1994).

These sanguine results notwithstanding, we believe that on balance our analysis is consistent with the conclusions of Farley et al. (1994), Galster (1988), Massey and Denton (1993), Yinger (1986), and others who argue that both individual and institutional discrimination still contribute to the high levels of racial residential segregation. Negative stereotypes of blacks and Latinos are fairly common. Among whites these stereotypes and racial prejudice as a sense of group position translate into attitudes on residential integration. Among minorities, hostile feelings toward other groups shape resistance to residential integration. What is more, attitudes on residential integration appear to have a connection to actual places of residence. Each of these patterns is consistent with the mounting evidence of ongoing patterns of individual and institutional discrimination in the housing market (Turner 1992; Yinger 1991).

In particular, our results suggest that the high rates of black-white residential separation are likely to continue, even in an extraordinarily diverse metropolitan area such as Los Angeles. To be sure, Farley and Frey (1994) found that areas with larger nonblack minority populations had lower overall levels of black-white segregation in 1990. *Such communities, however, did not experience*

higher rates of decline in black-white segregation between 1980 and 1990. Our data may shed some light on this failure of the "buffer" hypothesis. A nontrivial fraction of Latinos and an even larger fraction of the Asian population hold negative stereotypes of blacks and react against the prospect of sharing residential space with blacks. Indeed, Asians were more hostile to blacks than were whites on each of our prejudice measures and in expression of residential integration attitudes. Hence, rather than operating as a "buffer" or source of greater options for blacks, ethnic diversity may in many instances simply add to the climate of resistance to blacks as neighbors.

Notes

1. It is necessary to note that we regard racial and ethnic distinctions as socially defined constructs (See & Wilson 1989; Stone 1985). Racial/ethnic categories and labels are variable across time and place in meaning as well as salience. Racial/ethnic distinctions also interact with class and gender. In designing the questionnaire, we decided to use the labels "whites," "blacks," "Asians," or "Asian Americans," and "Hispanic" or "Hispanic Americans." This decision rested on three considerations. First and foremost, our pretest results clearly indicated that these are widely used and well understood terms. Second, although it would have been ideal, it was not feasible to develop a questionnaire fully sensitive to even a limited range of national origin subgroups within the heterogeneous Latino (i.e., Mexican, Nicaraguan, Salvadoran, etc.) and Asian (i.e., Korean, Japanese, Chinese, etc.) populations. Third, there are strong reasons to believe that social interaction and discourse are often dominated by reliance upon stereotypes linked to the broad racial/ethnic categories we have used (see Bobo & Hutchings 1994 for a fuller discussion). This is not to deny the importance of variability based on national origin, class, and gender. Rather, it is a concession to the powerful evidence that racial/ethnic cues involving these four broad categories often precede and sometimes overwhelm these other group statuses in social interaction, especially where questions of racial residential integration are concerned.

2. Farley et al. (1993) show that Detroit area blacks made rent and mortgage payments comparable to those necessary for many suburban residences. Among homeowners, the black-white gap was $250; however, 60% of blacks owning homes in the central city area and 52% of suburban whites had monthly housing expenditures between $300 and $799 per month. The gap for renters was even smaller, and 61% of both blacks and whites had gross rent payments between $300 and $599 per month. Farley et al. (1993) point out that the Detroit area is characterized by a declining housing market, reporting a 6% decrease — to $68,000 — in the median price of single-family dwellings between 1980 and 1990.

3. Despite Clark's claims of neutral ethnocentrism, Clark finds that whites express the strongest preferences for having a majority of their neighbors from their own race, with almost no whites expressing interest in living in neighborhoods that are less than 50% white. Moreover, whites, Hispanics, and Asians all express the strongest desire for own-race neighbors when the target group is blacks, and blacks and Asians express the strongest desire for neighborhoods with a substantial number of whites (Clark 1992). Blacks' expressed preferences for 50/50 white neighborhoods may represent preferences for higher-quality neighborhoods, rather than a desire to live and interact with whites (Clark 1992:455). Neighborhoods with a substantial number of whites often have higher-quality schools and services, and blacks — and other excluded groups — are well aware of these differences. While there is likely to be some validity to the desire for higher-quality neighborhood services (Farley et al. 1993:22), blacks also place a great deal of importance on integration, racial harmony, and effective interaction with whites (Farley et al. 1978; Farley et al. 1993; Pettigrew 1973).

4. We compared the LACSS data to 1990 census distributions on nativity, sex, education, age, family income, and occupation. The only noteworthy differences emerged, generally, on the education variable and did so in ways typical for telephone surveys. Our sample is better educated than the population at large, particularly among blacks and Latinos. It should be

borne in mind that not all households have telephones in them and that telephone coverage varies by social class and by race/ethnicity, with whites more often living in telephone households than do blacks or Latinos (Groves & Kahn 1979; Thornberry & Massey 1988). In particular, the LACSS data underrepresented high school graduates. In addition, as a result of the large number of Asian nationality groups and languages — 19 different nations of origin were found among Asians in the sample — it was impractical to develop additional foreign language translations of the questionnaire. Thus, even though 70% of our Asian respondents were nonnatives, this figure is substantially below the 1990 census figures for Los Angeles County (88%).

5. The data are not shown, but can be obtained from the authors upon request.

6. The format of the stereotyping measures was taken directly from the 1990 General Social Survey (Bobo & Kluegel 1991; Smith 1991), which included a series of bipolar trait-rating items. The bipolar trait-rating format has been shown to increase respondent comfort in expressing racial/ethnic group distinctions (Jackman & Senter 1983), to produce evidence of much higher levels of negative stereotyping than do older forced-choice formats (e.g., compare reported perceptions of blacks as less intelligent from Schuman, Steeh & Bobo 1985 to those reported by Bobo & Kluegel 1991 and Smith 1991), and to yield reasonably reliable and valid measures of prejudice (Bobo & Kluegel 1993; Farley et al. 1994).

7. To this point our analyses strongly contradict the mere in-group preference hypothesis advanced by Clark. One possible objection to our results could be that we find a much stronger role for prejudice than for mere in-group preferences because of the dependent variable. Our residential integration measures tap out-group social distance, which differs from expression of preferred neighborhood composition. In order to see whether our results are robust using a preference dependent measure, we conducted analyses of data from the 1976 National Election Study Survey (Miller & Miller 1977). This national probability sample of adults 18 years of age or older contained feeling thermometer measures and a neighborhood preference question. The latter asked: "Would you personally prefer to live in a neighborhood that is all (white/black), mostly (white/black), about half (white/black) and half (black/white), or mostly (black/white)?" For the analyses reported below, high scores indicate responses of mostly opposite race and low scores indicate expressed preference for all same-race neighbors. Unfortunately, the 1976 NES did not include reasonable proxies for perceived economic status differences or group stereotypes.

As was found for our 1992 LACSS data, the thermometer measures for both black and white respondents reveal clear in-group preference or ethnocentrism. Blacks' mean rating for "blacks" was 86.7, as compared to 68.4 for "whites." Among whites, the mean in-group rating was 74.2 as compared to 58.8 for "blacks." Among blacks we found no significant correlation between in-group rating (r = -.08), out-group rating (r = .02), or affective differentiation (r = -.11) and responses to the neighborhood preference question. Among whites, there are significant correlations between neighborhood preference and in-group rating (r = -.23), out-group rating (r = .23), and, most important, affective differentiation (r = -.38). This pattern of associations among white respondents is consistent in the South and non-South.

OLS regression models support our earlier findings with one qualification. The strongest predictor of neighborhood preferences, net of education, age, region, sex, and family income among whites, was the affective differentiation measure. Entered singly, or jointly, however, we do find significant effects of in-group affect and of out-group affect among white NES respondents. Measures of in-group affect appear to play a stronger role when predicting a neighborhood composition preference item as opposed to a social distance item (setting aside for the moment possible change over time and sample composition differences between the 1976 NES and 1992 LACSS data). Among blacks, only the affective differentiation measure has a significant effect. On the whole, however, examination of a different dependent measure confirms key findings based on the LACSS data.

8. Patterns for the questions on perceived neighborhood racial composition parallel analysis of actual residential patterns from census data in two other important ways. First, our data are consistent with census data showing that middle class blacks are less able to translate their class status into greater residential proximity to whites (Farley & Frey 1994; Kain 1986). Eighty-one percent of blacks with family incomes of $60,000 or more reported living in mostly black neighborhoods as compared to just 11% who report living in mostly white neighborhoods. In

contrast, 33% of Asians in this high-income category reported living in largely Asian neighborhoods and an equal 33% report living in largely white areas (three times the rate for high-income blacks). Similarly, only 26% of high-income Hispanics reported living in largely Hispanic neighborhoods, and 47% reported living in largely white neighborhoods (more than four times the rate observed among blacks).

Second, among both Latinos and Asians, native respondents — more so than the foreign-born — report that they live in largely white neighborhoods (Ong, Lawrence & Davidson 1992). These figures are 19% versus 9% comparing native-born and foreign-born Latinos. The comparison for Asians is 24% versus 19%.

References

Allport, Gordon W. 1954. *The Nature of Prejudice.* Doubleday Anchor.

Berry, Brien J.L. 1979. *The Open Housing Question.* Ballinger.

Blumer, Herbert. 1958. "Race Prejudice as a Sense of Group Position." *Pacific Sociological Review* 1:3-7.

Bobo, Lawrence. 1988. "Attitudes toward the Black Political Movement: Trends, Meaning and Effects on Racial Policy Preferences." *Social Psychology Quarterly* 51:287-302.

____. 1989. "Keeping the Linchpin in Place: Testing the Multiple Sources of Opposition to Residential Integration." *International Review of Social Psychology* 2:305-23.

Bobo, Lawrence, and Vincent L. Hutchings. 1994. "Black and Latino Conflict with Asians: Extending the Theory of Group Position." Paper presented at the meetings of the Western Political Science Association, Albuquerque, New Mexico.

Bobo, Lawrence, James H. Johnson Jr., Melvin L. Oliver, Jim Sidanius, and Camille L. Zubrinsky. 1992. *Public Opinion before and after* A Spring of Discontent: *A Preliminary Report on the 1992 Los Angeles County Social Survey.* Occasional Working Paper Series. Vol. 3, no. 1. Center for the Study of Urban Poverty.

Bobo, Lawrence, and James R. Kluegel. 1991. "Modern American Prejudice: Stereotypes, Social Distance, and Perceptions of Discrimination toward Blacks, Hispanics, and Asians." Paper presented at the annual meeting of the American Sociological Association, Cincinnati, Ohio.

____. 1993. "Opposition to Race-Targeting: Self-Interest, Stratification Ideology, or Racial Attitudes?" *American Sociological Review* 58:443-64.

Bobo, Lawrence, Howard Schuman, and Charlotte Steeh. 1986. "Changing Attitudes toward Residential Integration." Pp. 152-69 in *Housing Desegregation and Federal Policy,* edited by John M. Goering. University of North Carolina Press.

Cell, John. 1981. *The Highest Stage of White Supremacy.* Oxford University Press.

Clark, W.A.V. 1986. "Residential Segregation in American Cities: A Review and Interpretation." *Population Research and Policy Review* 5:95-127.

____. 1988. "Understanding Residential Segregation in American Cities: Interpreting the Evidence, a Reply to Galster." *Population Research and Policy Review* 7:113-21.

____. 1991. "Residential Preferences and Neighborhood Racial Segregation: A Test of the Schelling Segregation Model." *Demography* 28:1-19.

____. 1992. "Residential Preferences and Residential Choices in a Multiethnic Context." *Demography* 29:451-66.

Farley, Reynolds, and William H. Frey. 1994. "Changes in the Segregation of Whites from Blacks during the 1980s: Small Steps toward a More Integrated Society." *American Sociological Review* 59:23-45.

Farley, Reynolds, Charlotte Steeh, Tara Jackson, Maria Krysan, and Keith Reeves. 1993. "Continued Racial Residential Segregation in Detroit: 'Chocolate City, Vanilla Suburbs' Revisited." *Journal of Housing Research* 4:1-38.

Farley, Reynolds, Charlotte Steeh, Maria Krysan, Tara Jackson, and Keith Reeves. 1994. "Stereotypes and Segregation: Neighborhoods in the Detroit Area." *American Journal of Sociology* 100:750-80.

Farley, Reynolds, Howard Schuman, Suzanne Bianchi, Diane Colasanto, and Shirley Hatchett. 1978. "'Chocolate City, Vanilla Suburbs': Will the Trend toward Racially Separate Communities Continue?" *Social Science Research* 7:319-44.

Galster, George. 1988. "Residential Segregation in American Cities: A Contrary Review." *Population Research and Policy Review* 7:93-112.

____. 1989. "Residential Segregation in American Cities: A Further Response to Clark." *Population Research and Policy Review* 8:181-92.

Greenwell, Michael, Mark Strohm, and Lawrence Bobo. 1994. *Project Memoranda: Non-Response Bias Evaluation.* Department of Sociology, University of California, Los Angeles.

Groves, Robert M., and Robert L. Kahn. 1979. *Surveys by Telephone: A National Comparison with Personal Interviews.* Academic Press.

Hirschman, Charles. 1983. "America's Melting Pot Reconsidered." Pp. 397-423 in *Annual Review of Sociology.* Vol. 9. Annual Reviews.

Institute for Social Science Research. 1992. *Los Angeles County Social Survey.* University of California, Los Angeles, Survey Research Center.

Jackman, Mary R. 1977. "Prejudice, Tolerance, and Attitudes toward Ethnic Groups." *Social Science Research* 6:145-69.

____. 1994. *The Velvet Glove: Paternalism and Conflict in Gender, Class, and Race Relations.* University of California Press.

Jackman, Mary R., and Marie Crane. 1986. "'Some of My Best Friends Are Black . . .': Interracial Friendship and Whites' Racial Attitudes." *Public Opinion Quarterly* 50:569-82.

Jackman, Mary R., and Robert W. Jackman. 1983. *Class Awareness in the United States.* University of California Press.

Jackman, Mary R., and Michael J. Muha. 1984. "Education and Intergroup Attitudes: Moral Enlightenment, Superficial Democratic Commitment, or Ideological Refinement?" *American Sociological Review* 49:751-69.

Jackman, Mary R., and Mary Scheuer Senter. 1983. "Different Therefore Unequal: Beliefs about Groups of Unequal Status." *Research in Social Stratification and Mobility* 2:309-35.

Jaynes, Gerald David, and Robin M. Williams Jr. 1989. *A Common Destiny: Blacks and American Society.* National Academy Press.

Kain, John F. 1986. "The Influence of Race and Income on Racial Segregation and Housing Policy." Pp. 99-118 in *Housing Desegregation and Federal Policy,* edited by John M. Goering. University of North Carolina Press.

Katz, Irwin. 1991. "Gordon Allport's *The Nature of Prejudice.*" *Political Psychology* 12:125-57.

Leven, Charles L., James T. Little, Hugh O. Nourse, and R. Read. 1976. *Neighborhood Change: Lessons in the Dynamics of Urban Decay.* Ballinger.

Lieberson, Stanley. 1980. *A Piece of the Pie: Blacks and White Immigrants since 1880.* University of California Press.

Massey, Douglas S., and Nancy A. Denton. 1989. "Hypersegregation in U.S. Metropolitan Areas: Black and Hispanic Segregation along Five Dimensions." *Demography* 26:373-92.

____. 1993. *American Apartheid.* Harvard University Press.

Massey, Douglas S., and Andrew B. Gross. 1991. "Explaining Trends in Racial Segregation, 1970-1980." *Urban Affairs Quarterly* 27:13-35.

Miller, Warren E., and Arthur H. Miller. 1977. *The CPS 1976 American National Election Study Introduction and Codebook.* Inter-University Consortium for Political and Social Research.

Muth, Richard F. 1986. "The Causes of Housing Segregation." *Issues in Housing Discrimination: A Consultation/Hearing of the United States Commission on Civil Rights.* Washington, D.C., 12-13 November, 1985, vol. 1: 3-13.

O'Neil, Michael J. 1979. "Estimating the Non-Response Bias Due to Refusals in Telephone Surveys." *Public Opinion Quarterly* 43:218-32.

Ong, Paul M., and Janette R. Lawrence, with Kevin Davidson. 1992. "Pluralism and Residential Patterns in Los Angeles." Unpublished paper, School of Architecture and Urban Planning, University of California, Los Angeles.

Pearce, Diana M. 1979. "Gatekeepers and Homeseekers: Institutional Patterns in Racial Steering." *Social Problems* 26:325-42.

Pettigrew, Thomas F. 1973. "Attitudes on Race and Housing: A Social Psychological View." Pp. 43-58 in *Segregation in Residential Areas,* edited by Amos H. Hawley and V.P. Rock. National Academy of Science Press.

____. 1979. "Racial Change and Social Policy." *The Annals of the American Academy of Political and Social Science* 441:114-31.

____. 1982. "Prejudice." Pp. 1-29 in *Dimensions of Ethnicity: Prejudice,* edited by S. Thernstrom, A. Orlov, and O. Handlin. Belknap.

Schuman, Howard, and Lawrence Bobo. 1988. "Survey-Based Experiments on White Racial Attitudes toward Residential Integration." *American Journal of Sociology* 94:273-99.

Schuman, Howard, Charlotte Steeh, and Lawrence Bobo. 1985. *Racial Attitudes in America: Trends and Interpretations.* Harvard University Press.

Sears, David O. 1988. "Symbolic Racism." Pp. 53-84 in *Eliminating Racism: Profiles in Controversy,* edited by Phyllis A. Katz and Dalmas A. Taylor. Plenum.

See, Katherine O'Sullivan, and William Julius Wilson. 1989. "Race and Ethnicity." Pp. 223-42 in *Handbook of Sociology,* edited by Neil J. Smelser. Sage.

Sigelman, Lee, and Susan Welch. 1993. "The Contact Hypothesis Revisited: Black-White Interaction and Positive Racial Attitudes." *Social Forces* 71:781-95.

Smith, Tom W. 1991. "Ethnic Images." GSS Technical Report no. 19. National Opinion Research Center.

Stone, Jonathan. 1985. *Racial Conflict in Contemporary Society.* Harvard University Press.

Thornberry, Owen T., and James T. Massey. 1988. "Trends in United States Telephone Coverage across Time and Subgroups." Pp. 25-49 in *Telephone Survey Methodology,* edited by Robert M. Groves et al. Wiley.

Turner, Margery Austin. 1992. "Discrimination in Urban Housing Markets: Lessons from Fair Housing Audits." *Housing Policy Debate* 3:185-215.

Waldinger, Roger. 1989. "Immigration and Urban Change." *Annual Review of Sociology* 15:211-32.

Woodward, C. Vann. 1974. *The Strange Career of Jim Crow.* Oxford University Press.

Yinger, John. 1986. "Measuring Racial Discrimination with Fair Housing Audits: Caught in the Act." *American Economic Review* 76:881-93.

____. 1991. *Housing Discrimination Study: Incidence and Variation in Discriminatory Behavior.* U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

The Socioeconomic Assimilation of Caribbean American Blacks*

MATTHIJS KALMIJN, *Utrecht University*

Abstract

Several authors have examined whether black Caribbean immigrants are more successful in the American economy than African Americans. This study examines the earnings and occupations of Caribbean American men in the 1990 census and expands previous analyses by examining generational differences within this new black minority. Central findings suggest that: (1) There continue to be important socioeconomic differences between Caribbean American blacks and African Americans. (2) If we can speak of a "black success story," however, it is the story of British Caribbeans; blacks from the French- and Spanish-speaking Caribbean do worse than African Americans. (3) Second- and later-generation Caribbean blacks generally have higher socioeconomic status than the immigrant generation. (4) For British Caribbeans, this implies that later generations have gained further advantages on African Americans. The pattern of generational differences within the Caribbean American community generally does not support a notion of structural assimilation into the black American community.

Although Caribbean American blacks form a small group in American society, they are important from a sociological point of view in that they have a dual status, a status as "black" and a status as "immigrant." The dual status they face raises important questions about the ethnic identity they form of themselves as a group (Waters 1991), the social and political boundaries they maintain vis-á-vis native-born blacks (Kasinitz 1992), and the way they are treated in labor markets and other public spheres by the white American majority (Model 1991). One of the more intensely debated issues in this respect is the question of whether Caribbean blacks are more successful in the American economy than native-born blacks. Some authors regard Caribbean Americans as a "black success story" in a racially segregated society (Sowell 1978; Glazer & Moynihan 1964), while others believe they face the double hurdle of xenophobia and racism in the American labor market (Bryce-Laporte 1972).

The assimilation of Caribbean blacks in the U.S. has been studied as far back as the 1930s (Reid 1939), but only recently have researchers begun to

** An earlier version of this article was presented at the annual meeting of the American Sociological Association in Los Angeles in 1994. I thank Frank Dobbin for encouraging comments and helpful suggestions. Address correspondence to Department of Sociology, Utrecht University, Heidelberglaan 1, Utrecht 3584 CS, Netherlands.*

examine their socioeconomic status in a systematic fashion. In an analysis of the 1970 census, Sowell (1978) found that first- and second-generation West Indian blacks had higher family incomes and were more likely to be employed in the professions than native-born blacks. Differences were found both in the U.S. as a whole, and in the New York metropolitan area, where most West Indian blacks live. Although Sowell did not take into account differences in background characteristics, multivariate analyses of the 1970 census by Chiswick (1979) led to similar conclusions. Chiswick compared foreign-born blacks, most of whom are from the West Indies, to native-born blacks and showed that ten years after arrival, foreign-born blacks had higher annual earnings than native-born blacks, even after the influence of education and experience was taken into account. While these findings are consistent with the stereotype of Caribbeans as a black model of achievement, analyses of the 1980 census suggest that the gap between Caribbean and African Americans has narrowed. In a multivariate analyses of Caribbean immigrants in the 1980 census, Butcher (1994) shows that after controlling for education, experience, and several other background characteristics, Caribbean immigrants were more likely to be employed than African Americans. Conditional on employment, however, they did not have higher annual or weekly earnings than native-born blacks. In a comparable analysis of the 1980 census, Model (1991) found the same result for foreign-born blacks in states with large Caribbean populations (e.g., New York, Florida, Massachusetts). In their monograph of the 1980 census, Farley & Allen (1989) found no differences in earnings either, but they also showed that foreign-born blacks were more often employed in professional and managerial occupations than native-born blacks. While comparisons of studies with different measures and models should be made with care, previous analyses generally lead to the conclusion that if there is a Caribbean advantage, it was greater in 1970 than in 1980, and in 1980, it seemed limited to advantages in the occupational domain.

Using 1990 census data on earnings and occupation, I examine generational differences within this new black minority. Assimilation theory argues that through a process of generational replacement, national origin groups gradually become more similar to the members of the host society (Hirschman 1983). For Caribbean Americans, the assimilation process is more uncertain because if they are losing part of their cultural heritage, the possibility arises that they are identified with the larger black American community (Foner 1985). Cultural assimilation in this context may imply that members of the second generation lose some of the advantages that their immigrant parents had over African Americans, and that their race again becomes a handicap in the status attainment process. In this study, I examine the socioeconomic aspect of the Caribbean American assimilation process by comparing the earnings and occupations of American-born Caribbean men with those of Caribbean immigrant and African American men.

Next to examining generational differences, I contribute to previous research in three ways. First, I update findings of 1970 and 1980 census analyses. Recent data are important because the earnings of native American blacks have deteriorated considerably during the 1980s (Bound & Freeman 1989). The question thus arises whether Caribbean blacks have suffered to the same extent as native blacks, or whether they have been protected from the economic

downturn of the 1980s. Second, I emphasize the heterogeneity of the Caribbean community by examining socioeconomic differences with respect to linguistic group (i.e., Spanish, French, and British-speaking Caribbeans). Does the "black success story" apply to all Caribbean groups, or is it limited, as several authors suggest, to those of British Caribbean descent? Third, I systematically compare differences in occupation and earnings, thus broadening the scope of previous analyses. Are socioeconomic differences in 1990 limited to the occupational domain, or do they also apply to advantages in earnings?

Background

In the past, several authors have tried to explain the socioeconomic success of Caribbean immigrants in the U.S. (Bryce-Laporte 1972; Butcher 1994; Foner 1985; Model 1991; Sowell 1978). These explanations can be grouped in hypotheses focusing on characteristics of the immigrants themselves (supply-side hypotheses), and hypotheses focusing on the way Caribbeans are treated in the American labor market (demand-side hypotheses).

Hypotheses focusing on the supply-side of the labor market argue that Caribbean blacks have a greater motivation for achievement and a stronger work ethic than native-born blacks. Originally, this argument was proposed by Sowell (1978), who attributed differences in work ethic to differences between the American and Caribbean systems of slavery. In a frequently contended essay, Sowell (1978) argued that because slaves in the Caribbean did not experience strong economic competition from a large white lower class, as slaves did in the American south, they had more opportunity for economic initiative and were socialized into a "spirit of capitalism" early on (46). Sowell's argument has been considered provocative because it seemed to imply that a "lack of achievement" orientation, rather than racism or discrimination on the basis of color, is the cause of racial inequality in American society. When we compare the position of West Indian blacks to native whites, however, this implication is disputable. Even if one assumes a greater work ethic on the part of West Indian blacks, one still needs another theory to explain why this has not brought them on a par with whites. Since the West Indian disadvantage vis-á-vis whites is large, Sowell's conclusion that "the West Indian success pattern undermines the explanatory power of current white discrimination as a cause of current black poverty" (49) is exaggerated. In addition, his argument that the position of West Indians in the contemporary U.S. can be traced to differences between the economies of the American South and the Caribbean many generations ago is difficult to prove or to refute, and hence less practical as an explanation.

A more obvious hypothesis about the motivation and work ethic of Caribbean immigrants focuses on differences between American and Caribbean society in modern times. Although many American blacks have entered the middle class in the last decades, structural changes in American society, such as the deteriorating manufacturing industry and the flight of the middle class to the suburbs, have deprived many young urban blacks of successful models for achievement in their community (Wilson 1987). Although most of the Antilles

face serious economic problems as well, blacks have traditionally formed the majority in the Caribbean islands and have occupied a more diverse range of positions in the social hierarchy. The more positive examples of black achievement and the less rigid social classification of racial groups in these societies may well lead to a stronger orientation toward socioeconomic achievement among Caribbean blacks. This argument would apply more to the British Caribbean, where blacks form the majority, than to the Hispanic Caribbean, where they generally constitute a minority.

A second reason why Caribbean immigrants may have a stronger motivation for achievement lies in the process of migration itself. Since people who voluntarily come to the U.S. tend to migrate for economic reasons, they may be specially motivated to perform well in the American economy and may have more talents and abilities that are relevant for socioeconomic achievement than people who do not make the move abroad. Since economically motivated migration is considered to be more selective with respect to economic skills and abilities than politically driven migration (Chiswick 1979), this hypothesis is particularly plausible for immigrants from the English-speaking Caribbean, almost all of whom were motivated by the desire to escape poor domestic economies (Palmer 1974). Because individual data comparing who migrates and who stays behind are scarce, it has been difficult to prove the selective migration hypothesis directly. Indirect evidence for the Caribbean context comes from an old and small-scale study by Tidrick (1971), which shows that students in two Jamaican universities who planned to move to the U.S. were more competitive and more strongly oriented toward economic success than those who planned to stay home, even after considering the influence of class background.

Next to hypotheses focusing on characteristics of the immigrants themselves are arguments focusing on the demand side of the labor market. It has long been argued that the main reason why Caribbean blacks are more successful in the American economy than native born blacks is that they suffer less from racial discrimination by whites (Reid 1939). Some assume that they are favored because of their British accents; others believe they present themselves to prospective or current employers as culturally distinct from, and possibly superior to, native American blacks. These arguments have not been proven systematically, but anecdotal evidence confirms that Caribbeans are perceived as more favorable (Foner 1985). The argument of white favoritism applies more to the English- and French-speaking Caribbean than to the Spanish-speaking Caribbean because the latter may face the risk of double discrimination as blacks and Hispanics in the American labor market.

My study elaborates on earlier analyses of nativity differentials within the black community by comparing Caribbean immigrants to American born blacks with Caribbean ancestry. Both supply- and demand-side hypotheses suggest that second-generation Caribbean blacks would be less successful in terms of achievement than Caribbean immigrants. Second-generation Caribbeans have not experienced a more favorable climate of race relations in the Caribbean, and hence may lack part of the self-confidence that Caribbean immigrants presumably bring with them. In addition, members of the second generation are not selected in the migration process, though they may display some of their

parents' traits. The second generation is also less likely to return if achievements in the U.S. fail to meet expectations.

Demand-side hypotheses imply similar outcomes. The children and grandchildren of Caribbean immigrants have not gone to schools that are heavily based on British or other European traditions, they are probably less likely to be exposed to Caribbean culture, and they may be more likely to speak English fluently or without a British accent. If white employers are less able to distinguish second-generation Caribbean blacks from African Americans, American-born Caribbeans will have less of an advantage over African Americans than Caribbean immigrants.

In summary, an elementary assimilation perspective implies that the children of Caribbean immigrants are becoming more similar to African Americans than their parents. Caribbean blacks thus represent a case where cultural assimilation may hamper rather than improve their socioeconomic achievement. While the scenario of cultural assimilation is plausible in light of what many white national origin groups have experienced in American society (Hirschman 1983), the question of whether Caribbeans lose the cultural distinctiveness that characterized their immigrant parents remains open (Foner 1985). In this study, I address this question by examining the socioeconomic dimension of the Caribbean American assimilation process. I describe income and occupational differences between Caribbean and African Americans, between different generations of Caribbeans, and between different linguistic groups within the Caribbean community. Though census data provide important information on such patterns, they do not allow me to decide between the demand- and supply-side hypotheses directly.

Data

The data I use are the Public Use Microsamples of the 1990 census (U.S. Bureau of the Census 1992). Using the census questions on a person's ancestry and country of birth, I define Caribbean Americans as black immigrants who were born in the Caribbean (first-generation Caribbeans), and American-born blacks who reported Caribbean ancestry (second- and later-generation Caribbeans). African Americans are defined as American-born blacks who did not report a specific ancestry. Of this latter group, 85% answered "black" or "African American" in response to the ancestry question, 3% answered "American," and 12% did not answer the question. That few native-born African Americans report a specific ancestry is plausible given the fact that the vast majority of them have been here for several generations. Because the census dropped the question on foreign parentage after the 1970 census, ancestry data are the only possible way of identifying second and later generation Caribbeans in recent censuses. Lieberson & Santi (1985) have compared classifications on the basis of ancestry and foreign parentage and concluded that inferences regarding educational and economic characteristics were generally similar under the two systems.

African Americans are obtained from a 25% sample of individual records in the 1% microsample. Because of their small numbers, Caribbean Americans are

obtained from the full 5% microsample. The samples I analyze are limited to the noninstitutionalized population, they exclude persons who were born abroad of American parents, and they do not include persons in the U.S. outlying areas. Except for the analyses in Figures 1 and 2, which show the geographical origins and destinations of Caribbean blacks, all my analyses focus on men. For a comparison with women, see Model (1991).

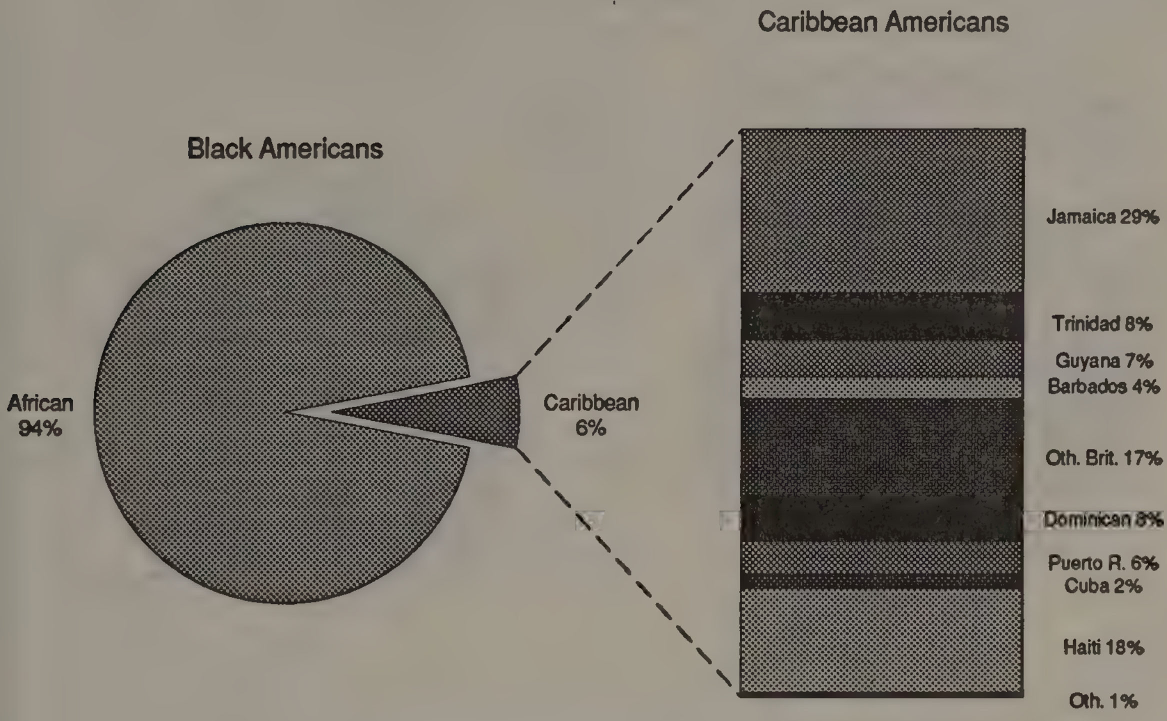
As Figure 1 shows, about 6% of the black community 16 years or older is Caribbean American. Caribbean blacks come from a large number of islands, but a few countries make up the bulk of the immigration flow: Jamaica (29%), Haiti (18%), the Dominican Republic (8%), and Trinidad-Tobago (8%). Considering differences with respect to immigration history and colonial heritage (Reid 1939), the various national origin groups can be classified in three linguistic groups: the English-speaking West Indians (65%), the Spanish-speaking West Indians (16%), and the French-speaking West Indians (19%). Persons are classified based on where they came from or where their ancestors came from, not on the language they themselves speak (which is affected by the assimilation process). The English-speaking Caribbean countries included are Jamaica, Trinidad-Tobago, Guyana, Barbados, Belize, the Bahamas, the Virgin Islands, and a range of smaller islands such as the Cayman and Leeward Islands. The main Spanish-speaking Caribbean countries are the Dominican Republic, Puerto Rico, and Cuba. The French-speaking Caribbean countries are Haiti, Guadeloupe, Martinique, and French Guyana. Note that the term *French West Indian* is often used to refer to Martinique and Guadeloupe, while the population analyzed here is almost exclusively from Haiti. The Dutch-speaking West Indians (Surinam, Curacao, Aruba, Bonaire) are too small of a group in the U.S. to include in the analyses.

As is well known, Caribbean blacks are heavily overrepresented in the Northeast (Figure 2) and concentrated in a few states. About half of them live in New York, 17% in Florida, 7% in New Jersey, and 4% in Massachusetts. More importantly, virtually all of them live in large cities (primarily New York City, Miami, Newark, and Boston). I therefore decided to limit my comparisons to urban blacks. Though uncommon in the literature, the focus on urban blacks is warranted because rural and urban areas have widely different income and opportunity structures. If the aim is to compare groups in fairly similar structural positions, it seems inappropriate to include rural native-born blacks in a comparison with such a highly urbanized ethnic minority.

Differences in Socioeconomic Status

Socioeconomic status is measured by two related indicators, occupational status and hourly earnings. Occupational status is measured by the 1980 version of Duncan's socioeconomic index (SEI) of occupations (Stevens & Cho 1985). Because the 1980 and 1990 occupational classification systems are virtually the same, there were no problems in assigning SEI scores to 1990 census respondents. Earnings are measured by a person's 1989 income from wages, salaries, and self-employment. Because my main interest is in status attainment, not in labor supply, I consider hourly earnings, using information on the number of

FIGURE 1: National Origin of American Blacks 16 Years Old or Older in 1990

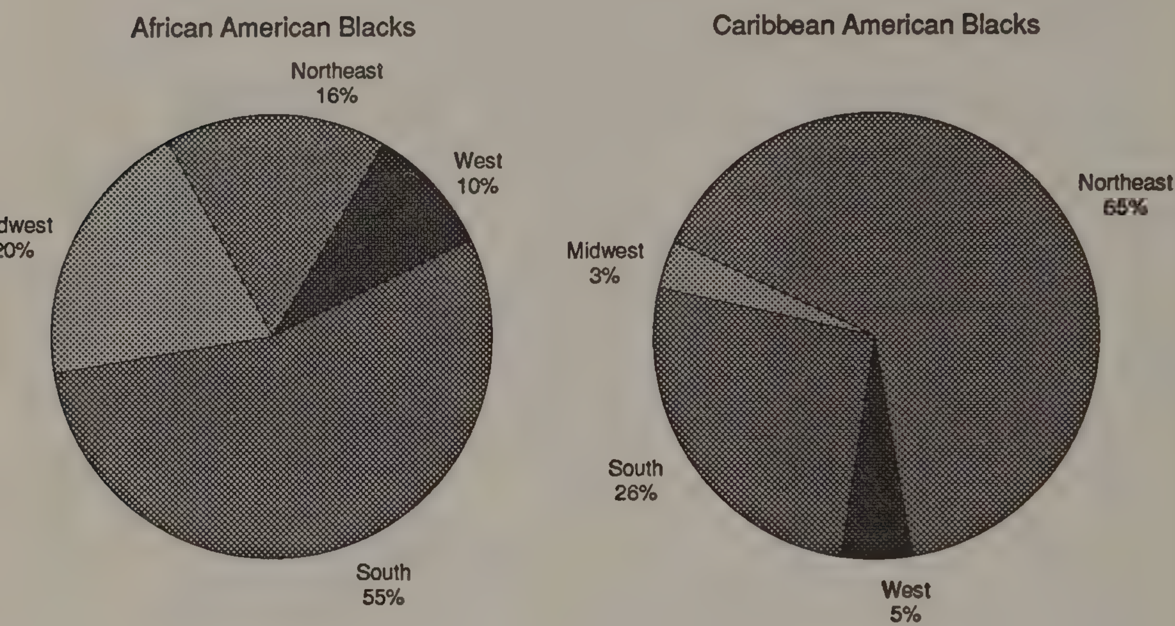


weeks worked in 1989 and the usual number of hours worked per week in that year. Only persons with positive annual earnings are considered.[1]

Since not all men work for pay, the possibility arises that my comparisons are affected by selection bias. Following previous research on black/white differences in earnings (e.g., Bound & Freeman 1989), I therefore limit my analyses and conclusions to a group for whom employment rates are high, that is, out-of-school men who were 25-64 years old at the time of the census. In addition, I use information on "last occupation" for men without a job at the time of the census. For persons who are not working, past occupation is probably the best predictor of the occupation they would have had, had they been in the labor force. Hence, the models of occupational attainment are practically unaffected by possible selection bias.

In the top panel of Table 1, I present data on mean earnings and occupational status for each of the four groups separately (African American, Hispanic Caribbean, French Caribbean, British Caribbean). Because extreme earnings may affect my comparisons, I also present mean logged earnings. Table 1 shows that British Caribbeans have higher earnings than African American blacks. On the other hand, men from the French- and Spanish-speaking Caribbean have lower earnings than African American blacks. It makes little difference here which mean we consider. Differences in occupational status parallel differences in hourly earnings. The British Caribbeans do best, followed by African Americans, Hispanic Caribbeans, and French Caribbeans respectively.

FIGURE 2: Geographical Location of Black Americans



Differences in Background Characteristics

The socioeconomic differences observed in Table 1 may be due to differences in human capital, differences in living arrangement, and differences in geographical location. I therefore estimate regression models to examine subgroup differences in socioeconomic status, net of the influence of such background characteristics. Table 1 (bottom panel) presents the means of the variables included in the regression models.

The first and foremost indicator of human capital is education. I include two measures of schooling: years of schooling completed and attainment of a college degree. Because the 1990 census question on schooling is not phrased in terms of years, as was the practice in earlier censuses, I approximate years of schooling by assigning the number of years it takes to receive the particular grade or degree.[2] I use the attainment of a college degree (B.A., M.A., Ph.D.) as an additional measure to capture possible wage or status premiums for college degrees on top of the benefits a person receives for additional years of schooling. Table 1 shows that on average, British Caribbeans complete slightly more years of schooling than African Americans. When focusing on the top end of the schooling distribution, the difference becomes clearer. The percentage of British Caribbeans who received a college degree is higher than the percentage of African Americans receiving a degree. Men from the Spanish- and French-speaking Caribbean, in contrast, are less educated than both African American and British Caribbean males.

A second human capital indicator is labor market experience and experience squared. Experience is approximated by age minus years of schooling minus 6.

TABLE 1: Means of Dependent and Independent Variables for Out-of-School Urban Black Males Aged 25-64 in 1990

	U.S.-born African American	Spanish-speaking Caribbean	French-speaking Caribbean	British-speaking Caribbean
Hourly earnings (in 1989)	12.41	11.56	11.26	13.64
Logged hourly earnings	2.28	2.16	2.16	2.39
Status of current or last occupation (SEI score)	30.2	28.3	27.7	33.2
Logged SEI score	3.27	3.22	3.18	3.37
Years of schooling	12.0	10.3	10.9	12.1
College degree (percent)	11.2	7.9	12.3	15.5
Years of experience	22.9	24.2	23.2	22.9
Poor English language ability (percent)	.4	31.6	23.0	.9
Married (percent)	49.4	55.5	60.0	64.4
Living in the West (percent)	12.1	4.4	1.8	6.9
Living in the Midwest (percent)	21.5	3.7	2.2	3.7
Living in the Northeast (percent)	17.9	76.6	50.0	63.9
Living in the South (percent)	48.5	15.3	46.0	25.5
Born in the U.S. (percent)	—	16.4	2.7	11.3
Years in U.S. (for immigrants)	—	17.0	12.5	14.7
Unweighted N	8,981	1,977	2,660	7,848

Although this measure captures potential rather than actual experience, it is to be preferred to a model with age and age squared, which does not take into account variations in the duration of schooling (Mincer 1974). Only small differences in experience are observed in Table 1.

A third indicator of human capital is English language ability. Table 1 shows that 32% of Hispanic and 23% of French Caribbeans rate their ability to speak English as poor or nonexistent. In the regression models, I include a dummy variable that takes the value of 1 for those whose language ability is poor (0 otherwise). Because virtually all British Caribbean and African American men have good English language ability, this specification implies that subgroup differences in socioeconomic status apply to men whose English is good.

I further include marital status because earlier studies have found that married men have higher earnings than single men, net of the influence of education and experience. Though part of this difference may be due to selection on the marriage market — highly paid men being more likely to marry — recent empirical studies have suggested that employer favoritism and greater productivity among married men are more important causes of the marital pay advantage (Korenman & Neumark 1991). Table 1 shows that all Caribbean American groups are more likely to be married than African American men. Since this gives them a socioeconomic advantage over African American men, marital status needs to be taken into account in the regression models.

Because a disproportionate number of Caribbean blacks live in New York, New Jersey, Florida, and Massachusetts, and since incomes are generally higher in these areas than, for instance, in the south, where the majority of African Americans lives (Figure 2), I take into account geographical differences by including a set of 4-1 binary variables referring to the region in which a person lives at the time of the census. To further check whether my findings are sensitive to regional differences, I also estimate models for specific regions separately.

Regression Models

CARIBBEAN VERSUS AFRICAN AMERICANS

In Table 2, I present ordinary least squares (OLS) estimates of unstandardized regression coefficients for hourly earnings and occupational status. The first model I consider is defined as follows,

$$\ln Y_i = \beta_0 + \gamma_1 Hisp_i + \gamma_2 Fren_i + \gamma_3 Brit_i + \epsilon_i \qquad \text{(A)}$$

In model A, $Hisp_i$, $Fren_i$, and $Brit_i$ are binary variables that are coded 1 if the individual is in the particular group (0 otherwise). Hence, the effects of these variables can be interpreted as the difference between Caribbean and African American blacks. The variable Y_i refers to occupational status and hourly earnings. Both measures are transformed into natural logarithms to reduce the influence of heteroskedasticity. Although transforming SEI scores to logarithms is uncommon, it facilitates interpretation of the coefficients and allows me to compare effects of a given variable across equations. Both can be regarded as percentage change in the dependent variable.[3]

Model A in Table 2 shows that there are large socioeconomic differences between Caribbeans and African Americans, though the magnitude and direction of these differences depend on which linguistic group we consider. The British Caribbean have a clear advantage over African Americans. For occupational attainment, the coefficient is .089, meaning that they have 9% higher occupational status than African Americans. For hourly earnings, the coefficient is .111, meaning that they have 12% higher earnings. Comparisons with the other two Caribbean groups yields the opposite picture. Compared to African Americans, the French-speaking Caribbean have 5% lower occupational status and 11% lower earnings. The Spanish-speaking Caribbean have 10% lower earnings and equal occupational status. Since no other variables are

TABLE 2: Regression of Occupational Status and Hourly Earnings on Selected Characteristics for Out-of-School Urban Black Males Aged 25-64 in 1990 (Model A and B)[a]

	SEI Score (Logged)		Hourly Earnings (Logged)	
	A	B	A	B
Independent variable				
Years of schooling	—	.038**	—	.038**
		(.001)		(.002)
College degree (0/1)	—	.480**	—	.297**
		(.012)		(.019)
Experience (in decades)	—	-.002	—	.249**
		(.011)		(.018)
Experience^2	—	.000	—	-.028**
		(.002)		(.004)
Poor English ability (0/1)	—	-.080**	—	-.167**
		(.020)		(.033)
Married (0/1)	—	.070**	—	.166**
		(.006)		(.010)
West versus South (0/1)	—	.101**	—	.174**
		(.010)		(.016)
Midwest versus South (0/1)	—	.009	—	.132**
		(.009)		(.014)
Northeast versus South (0/1)	—	.053**	—	.215**
		(.008)		(.013)
Spanish-speaking Caribbean (γ_1)[b]	-.096**	.021	-.114**	-.103**
	(.016)	(.018)	(.028)	(.029)
French-speaking Caribbean (γ_2)[b]	-.056**	-.042**	-.119**	-.089**
	(.019)	(.015)	(.024)	(.024)
British-speaking Caribbean (γ_3)[b]	.089**	.049**	.111**	.011
	(.010)	(.009)	(.014)	(.015)
Intercept	5.571**	5.014**	2.275**	1.226**
	(.004)	(.021)	(.006)	(.034)
R^2	.007	.271	.006	.134
F test comparing models B and A		813.8**		302.7**
Unweighted N	20,182	20,182	18,436	18,436

[a] Estimated standard errors in parentheses.
[b] Native-born African Americans are the reference group.

* $p < .05$ ** $p < .01$

included in the model, these results essentially replicate the differences in mean logged earnings observed in Table 1.

To examine whether these differences are due to the influence of human capital, living arrangements, and geographical location, I add a set of $j = 10$ control variables for each individual i,

$$\ln Y_i = \beta_0 + \sum_j \beta_j X_{ij} + \gamma_1 Hisp_i + \gamma_2 Fren_i + \gamma_3 Brit_i + \epsilon_i \quad \text{(B)}$$

In this model, the effects of $Hisp_i$, $Fren_i$, and $Brit_i$ can be interpreted as the difference between Caribbean and African American blacks when holding all other relevant measured characteristics constant.

Model B in Table 2 shows that for British Caribbeans, differences in socioeconomic status net of the influence of background characteristics are smaller than the differences observed in model A. Their occupational advantage over African Americans is reduced to 5% and their earnings advantage is reduced to 1%. When we compare total (model A) and net relative differences (model B), we can conclude that about half of the British occupational advantage and nearly the entire earnings advantage can be attributed to their more favorable background characteristics. That British Caribbeans have a net occupational advantage over African Americans without a corresponding earnings advantage suggests that they have difficulty in translating their occupational advantage into an earnings advantage. Perhaps this is due to a greater emphasis on — relatively poorly paying — high cultural segments of the labor market, such as teaching. That British West Indians have higher-status occupations but similar earnings is consistent with Farley & Allen's analysis of the 1980 census (1989).

For Hispanic and French Caribbeans, controlling for background characteristics changes the socioeconomic disadvantage they have compared to African Americans in a modest fashion. After controlling for the influence of background characteristics, Hispanic Caribbeans have 10% lower hourly earnings than African American males and 2% higher occupational status. Men from the French-speaking West Indies have both lower occupational status (4%) and lower earnings (9%).

Effects of the other independent variables are generally consistent with findings in previous studies of status attainment. Each additional year of schooling is associated with a 3.8% increase in occupational status and a 3.8% increase in hourly earnings. For obtaining a college degree, there is an occupational premium of 62% and an earnings premium of 35% in addition to the benefits of extra years of schooling. There is an increase in earnings for additional years of experience as well, but given the negative effect of the squared term, the rate of this increase declines over the working life.[4] A somewhat unexpected finding is that experience does not affect occupational status. Apparently, black men in the U.S. make monetary gains in careers without corresponding changes in occupation. Also, individuals whose English proficiency is poor or absent have 8% lower occupational status and 15% lower earnings. Married persons have higher earnings and occupational status. The

regional effects show that earnings and occupational status are highest in the Northeast and the West, followed by the Midwest and the South, respectively.

Although the model controls for geographical location, it remains possible that the socioeconomic gap between Caribbean and African American men is different in states where Caribbeans are concentrated. To examine this issue, I estimate model B for Caribbean states (i.e., New York, Massachusetts, and Florida) and non-Caribbean states separately. Subsequently, I test whether the relative socioeconomic advantages differ significantly across these two regions. The results of these comparisons, presented in Table 3, generally show that the relative position of Caribbean Americans is somewhat better in areas where they are not concentrated. The British Caribbean have more of an advantage over African Americans in the non-Caribbean states, and the Hispanic and French Caribbean have somewhat less of a disadvantage over African Americans in those states. These findings are consistent with earlier analyses showing that minority concentration is associated with greater gaps between minority and majority wages (Tienda & Lii 1987). Since these contrasts are not statistically significant, however, no firm conclusions can yet be made about the impact of minority concentration on immigrant earnings, and we may proceed the analysis by focusing on the Caribbean population at large.

GENERATIONAL DIFFERENCES

Are there generational differences within the black Caribbean community? To answer this question, we can compare Caribbean immigrants to native-born blacks of Caribbean ancestry (second and later generation Caribbeans). Comparing immigrants and natives is complicated by the fact that immigrants typically face some disadvantages in the labor market upon arrival in American society due to a lack of information on jobs and possibly a shortage of social capital to support the status attainment process as well. The longer they have been in the U.S., the more they overcome such problems, and at some point, they tend to reach socioeconomic parity with natives (Chiswick 1979; Chiswick & Sullivan 1995). Because assimilation may occur during an immigrant's lifetime, and not only between generations, we need to compare American born Caribbeans to Caribbean immigrants who have been in the U.S. for different periods of time.

To make these comparisons, I estimate the following regression model.

$$\begin{aligned}\ln Y_i = \beta_0 &+ \gamma_1 Hisp_i + \gamma_2 Fren_i + \gamma_3 Brit_i \\ &+ \gamma_4 Hisp_i x Imm_i + \gamma_5 Fren_i x Imm_i + \gamma_6 Brit_i x Imm_i \\ &+ \gamma_7 Hisp_i\, x\, Imm_i\, x\, Dur_i + \gamma_8 Fren_i\, x\, Imm_i\, x\, Dur_i + \gamma_9 Brit_i\, x\, Imm_i\, x\, Dur_i \\ &+ \epsilon_i\end{aligned} \tag{B}$$

In model B, Imm_i is coded 1 for immigrants (0 otherwise) and Dur_i is the number of years immigrants have been in the U.S., scaled in decades (0 for all other groups). Due to the specification of interaction effects chosen in model B, each parameter has a meaningful interpretation: γ_1 through γ_3 give the difference between native-born Caribbeans and African Americans; γ_4 through γ_6 give the difference between native-born Caribbeans and Caribbean immigrants upon arrival in the U.S.; γ_7 through γ_9 give the increase in earnings or

TABLE 3: Predicted Relative Differences between Caribbean and African American Males by Area of Residence (in Percentages Compared to African Americans)[a]

Socioeconomic indicator Origin group	Place of Residence: Entire U.S.	Caribbean[b] states	Non-Caribbean states	Difference (*t* test)
SEI score				
Spanish-speaking Caribbean	+2.1	-.7	+2.8	.9
French-speaking Caribbean	-4.1	-6.6	-.9	1.5
British-speaking Caribbean	+5.0	+3.1	+5.1	1.3
Hourly earnings				
Spanish-speaking Caribbean	-9.8	-11.7	-7.9	.7
French-speaking Caribbean	-8.5	-9.2	-4.8	.7
British-speaking Caribbean	+1.1	-.9	+4.5	1.7

[a] New York, Florida, and Massachusetts
[b] Net of the influence of other characteristics (see Table 2)

occupational status among immigrants for each ten years of duration in the U.S.[5] Estimates of model B appear in Table 3. For the sake of presentation, I do not report effects of the other independent variables in the equation.

Effects of duration of stay in a cross-sectional dataset of immigrants should be interpreted with some caution. Borjas (1985) has shown that the socioeconomic potential or "quality" of more recent immigrant cohorts is lower than that of immigrant cohorts who entered many decades ago, a phenomenon he attributes to changes in the 1965 Amendments to the 1952 Immigration and Nationality Act. Because duration of stay and year of immigration are by definition perfectly (negatively) correlated in a cross-section, assimilation during the life of an immigrant as inferred from cross-sectional analyses may in fact reflect a secular decline in the socioeconomic potential of entering immigrants (i.e., immigrant cohorts). There are several reasons, however, why an analysis of duration effects is still valuable in the present context. First, Borjas's argument applies more to the changing national origin mix of immigrants than to changes in the socioeconomic potential of specific national origin groups,

which is the issue examined here. Second, Borjas has shown that even when immigration cohort effects are removed, large effects of assimilation remain. Third, an important aspect of the socioeconomic potential of immigrants, i.e., educational attainment, is included in the model. It is somewhat difficult to imagine that there will be a decline in the socioeconomic potential across immigrant cohorts once educational attainment is controlled for. Fourth, in an analysis of cohort and duration effects for various national origin groups, Borjas (1987) has shown that for several of the sending countries considered here (e.g., Jamaica, Trinidad, Haiti, Cuba, with the exception of the Dominican Republic), there is no statistically significant decline in the socioeconomic potential of entering immigrants. Hence, the duration effects estimated in Table 4 will largely reflect assimilation during the life course of an immigrant.

The effects of the immigrant interactions in Table 4 are negative and generally statistically significant, indicating that natives have higher occupational status and earnings than immigrants who just arrived in the U.S. For example, in the occupational status equation, the effect for British immigrants is -.065, showing that immigrants have 6% lower occupational status than natives with similar amounts of education and experience. We further notice that all duration effects are positive, and generally statistically significant as well. Hence, immigrants gain status the longer they have been in the U.S. For British immigrants, the effect on occupational status is .028, indicating that they gain about 3% per decade. In general, we observe that assimilation occurs slower for British Caribbean immigrants than for the other Caribbean immigrants, probably because their human capital skills are more easily transferrable to the American labor market.

Although Caribbean natives are doing better than Caribbean immigrants who just arrived in the U.S., the question remains whether they are also doing better than immigrants who have been in the U.S. for longer periods of time. To answer this question, I use the model in Table 4 to make comparisons with immigrant Caribbeans at different periods of stay (5 years, 15 years, and 25 years).[6] These calculations show that even though Caribbean immigrants improve their position during their life in the U.S., at most of the relevant durations, they are still surpassed by native Caribbeans (Table 5). We can also compare native Caribbeans to the "average" Caribbean immigrant, that is, immigrants who have been in the U.S. for the average duration of the particular immigrant group. This comparison yields a similar pattern, though the generational advantage is clearly greater for French and Hispanic Caribbeans than for British Caribbeans. Hence, we can conclude that native Caribbeans have higher socioeconomic status than most immigrants, and that they are comparable to immigrants who have been in the U.S. for quite a long period of time. This pattern of assimilation is similar to what many white ethnic groups experienced in the U.S. (Hirschman 1983).

The final question is how immigrants and natives compare to the African American community. To answer this question we focus on native-born Caribbeans first. Table 4 shows that the main effects of language group (γ_1, γ_2, and γ_3) are generally positive, suggesting that native Caribbeans have higher earnings and occupations than African Americans. However, this effect is only significant for British Caribbeans, and for them, it is limited to the occupational

TABLE 4: Regression of Occupational Status and Hourly Earnings on Nativity and Duration in the U.S. for Out-of-School Urban Black Males Aged 25-64 in 1990 (Model C)[a]

Independent variable	SEI Score (Log)	Hourly Earnings (Log)
Native Caribbean versus African American		
Spanish-speaking Caribbean (γ_1)	.065	-.016
	(.038)	(.062)
French-speaking Caribbean (γ_2)	.058	.038
	(.087)	(.140)
British-speaking Caribbean (γ_3)	.070**	.000
	(.024)	(.038)
Immigrant versus native Caribbean		
Spanish x Immigrant (γ_4)	-.094*	-.249**
	(.050)	(.082)
French x Immigrant (γ_5)	-.205*	-.291*
	(.092)	(.148)
British x Immigrant (γ_6)	-.065*	-.062
	(.028)	(.045)
Duration effects for immigrants		
Spanish x Immigrant x Duration (γ_7)	.020	.077*
	(.016)	(.026)
French x Immigrant x Duration (γ_8)	.078**	.119**
	(.019)	(.031)
British x Immigrant x Duration (γ_9)	.028**	.051**
	(.009)	(.015)
R^2	.273	.136
F test comparing models C and B	5.03**	6.12**
Unweighted N	20,182	18,436

[a] Regression coefficients are net of the influence of other independent variables (listed in Table 2). Estimated standard errors in parentheses.

* $p < .05$ ** $p < .01$

TABLE 5: Predicted Relative Differences between Native and Immigrant Caribbean Black Males[a]

Socioeconomic indicator	Duration in the U.S./Immigrant Entry Cohort			
Origin group	5 years/ 1985	15 years/ 1975	25 years/ 1965	Average Immigrant[b]
SEI score				
Spanish-speaking Caribbean	+8.8	+6.6	+4.5	+6.2
French-speaking Caribbean	+18.1	+9.2	+1.0	+11.3
British-speaking Caribbean	+5.2	+2.3	-.1	+2.4
Hourly earnings				
Spanish-speaking Caribbean	+23.4	+14.3	+5.8	+12.5
French-speaking Caribbean	+26.0	+11.9	-.1	+15.3
British-speaking Caribbean	+3.7	-1.4	-6.3	-1.3

[a] In percentages compared to Caribbean immigrants
[b] Based on the models in Table 4.
[c] Predicted differences between natives and immigrants who have been in the U.S. for 17.0 years (Spanish-speaking Caribbeans), 12.5 years (for French-speaking Caribbeans), and 14.7 years (British-speaking Caribbeans).

domain. On average, British Caribbean natives have 7% higher occupational status than African Americans.

As shown before, immigrants have lower status than native Caribbeans, but they reach parity after a certain number of years in the U.S. At which point do Caribbean immigrants reach parity with African Americans? Simple algebraic manipulations of the equation in model B show that this point is reached when $Dur_i = -(\gamma_1 + \gamma_4)/\gamma_7$.[7] The point at which British Caribbeans immigrants begin to have an earnings advantage over African Americans is reached soon after they arrive (12 years). Their occupational advantage is already present when they arrive. For the other language groups, the picture is different. Immigrants from the Spanish- and French-speaking West Indies generally take much longer before they begin to have an advantage over African Americans (34 and 21 years for earnings, and 15 and 19 years for occupational status).

Conclusion

In the 1990s, there continue to be socioeconomic differences between Caribbean American blacks on the one hand, and African American urban men on the other but the results differ strongly by linguistic group. Both native British Caribbeans and British Caribbean immigrants who have been in the U.S. for more than a short period of time are doing better than African American blacks.

Though this result is consistent with the stereotype of Caribbeans as a "black success story" in American society, the British advantage is limited to the occupational domain and is not as spectacular in magnitude as is commonly believed. The advantage is not large enough, for example, to bring them on a par with white men (Bound & Freeman 1989). The picture is different for Caribbean blacks from the Spanish- and French-speaking West Indies. Here, immigrants are worse off than African Americans for most of their life in the U.S. After a long period of stay in the U.S., they eventually reach parity with their African Americans peers. Similarly, native-born Hispanic and French Caribbeans are not significantly different from the African American community.

That differences are language-specific is consistent with hypotheses about demand and supply factors. Supply-side hypotheses argue that Caribbeans have a greater motivation for achievement than native-born African Americans. Such differences can be attributed to selective migration on the one hand, and to socialization in black majority societies on the other. British Caribbeans tend to migrate for economic reasons, while Haitians migrate for political reasons. Despite the high social and emotional costs involved, political immigration or other forms of coerced immigration are generally believed to be less selective with respect to occupational ability than economically driven immigration. In addition, the British Caribbeans generally have black majorities while the French and the Hispanic West Indies considered here do not. Hence, the British may be more oriented towards achievement because they have been socialized in societies with black majorities and relatively successful economic role models among their peers. Support for demand-side hypotheses can be found as well. Hispanic West Indian blacks, for example, may face double discrimination because they are both black and Hispanic. Unlike the British, they do not combine their race with a more favorably looked upon British heritage.

Generational differences within this new black minority were found as well. On average, Caribbean blacks who were born in the U.S. have higher earnings and occupational status than Caribbean immigrants upon arrival. Immigrants reach parity with natives after several years in the U.S., but the point at which this occurs is in most cases well above the average number of years that Caribbean immigrants have been in the U.S. For British Caribbeans, this implies that the second and later generations have a somewhat greater advantage over African Americans than the first generation. This result does not support the notion suggested by assimilation theory that later generations lose the "immigrant" advantage that their parents had, and consequently reach parity with African Americans.

Several reasons can be suggested to explain this exception. First, cultural assimilation may be mitigated by the influence that immigrants have on their children. More specifically, Caribbean immigrants may transmit their orientation toward hard work and achievement by emphasizing the importance of schooling for their children. They may be particularly motivated to transmit such values if they contrast their own experiences in the U.S. with those of African Americans in their surroundings. Second, even though American-born Caribbean blacks were never faced with the decision of whether to migrate, the talents and abilities on which their parents are "selected" can be passed on to

later generations through the socialization process (Chiswick 1977). Third, intergenerational influences may be amplified and reinforced by geographical segregation between Caribbean and African Americans. Though there is some evidence pointing in this direction (Denton & Massey 1989), more research is needed to examine patterns of segregation between the two black communities.

Notes

1. Hourly earnings of $100 and more are recoded to 100.

2. The coding is done as follows: no school, nursery school and kindergarten (0 years), 1st-4th grade (2.5 years), 5th-8th grade (6.5 years), 9th grade (9 years), 10th grade (10 years), 11th grade (11 years), 12th grade and high school graduates (12 years), some college without degree and associate college degree in occupational program (13 years), associate college degree in academic program (14.5 years), Bachelor's degree (16 years), Master's degree (18 years), professional and doctorate degree (22 years).

3. For continuous independent variables, the percentage change in the dependent variable is given by $100 \times ß$; for binary independent variables, the percentage difference in the dependent variable between subgroups is given by $100 \times [\exp(ß)-1]$.

4. The point at which log earnings actually begin to decline with experience can be found by setting the derivative of equation B with respect to experience to 0 and rearranging terms. This yields $(-1 \times .249)/(2 \times -.028) = 4.4$ which amounts to 44 years of experience.

5. I also tested nonlinear duration effects using duration and duration squared but generally found no significant improvement of the models.

6. The model also allows one to compute at which point the socioeconomic status positions of immigrant and native Caribbeans are equal. For example, according to model B, the (logged) earnings of Hispanic Caribbean immigrants are: $ß_0 + \Sigma_j ß_j X_{ij} + \gamma_1 + \gamma_4 + \gamma_7 Dur_i$. The (logged) earnings of Hispanic Caribbean natives are: $ß_0 + \Sigma_j ß_j X_{ij} + \gamma_1$. Hence, the earnings of Hispanic immigrants and natives are equal when $Dur_i = -\gamma_4/\gamma_7$.

7. According to model B, the (log) earnings of Hispanic Caribbean immigrants are: $ß_0 + \Sigma_j ß_j X_{ij} + \gamma_1 + \gamma_4 + \gamma_7 Dur_i$. The (log) earnings of African Americans are: $ß_0 + \Sigma_j ß_j x_{ij}$. Hence, the earnings of Hispanic immigrants and African Americans are equal when $Dur_i = -(\gamma_1+\gamma_4)/\gamma_7$.

References

Borjas, George J. 1985. "Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants." *Journal of Labor Economics* 3:463-89.

_____. 1987. "Self-Selection and the Earnings of Immigrants." *American Economic Review* 77:531-53.

Bound, John, and Richard B. Freeman. 1989. "Black Economic Progress: Erosion of the Post-1965 Gains in the 1980s?" Pp. 32-49 in *The Question of Discrimination: Racial Inequality in the U.S. Labor Market*, edited by Steven Schulman and William Darrity Jr. Wesleyan University Press.

Bryce-Laporte, Roy Simon. 1972. "Black Immigrants: The Experience of Invisibility and Inequality." *Journal of Black Studies* 3:29-56.

Butcher, Kristin F. 1994. "Black Immigrants to the United States: A Comparison with Native Blacks and Other Immigrants." *Industrial and Labor Relations Review* 47:265-84.

Chiswick, Barry R. 1977. "Sons of Immigrants: Are They at an Earnings Disadvantage?" *American Economic Review* 67:376-80.

____. 1979. "The Economic Progress of Immigrants: Some Apparently Universal Patterns." Pp. 357-99 in *Contemporary Economic Problems*, edited by William Fellner. The American Enterprise Institute.

Chiswick, Barry R., and Teresa A. Sullivan. 1995. "The New Immigrants." Pp. 211-70 in *State of the Union: America in the 1990s. Volume Two: Social Trends*, edited by R. Farley. Russell Sage.

Denton, Nancy, and Douglas S. Massey. 1989. "Racial Identity among Caribbean Hispanics: The Effect of Double Minority Status on Residential Segregation." *American Sociological Review* 54:790-808.

Farley, Reynolds, and Walter R. Allen. 1989. *The Color Line and the Quality of Life in America*. Oxford University Press.

Foner, Nancy. 1985. "Race and Color: Jamaican Migrants in London and New York City." *International Migration Review* 19:708-27.

Glazer, Nathan, and Daniel Patrick Moynihan. 1964. *Beyond the Melting Pot*. M.I.T. Press and Harvard University Press.

Hirschman, Charles. 1983. "America's Melting Pot Reconsidered." *Annual Review of Sociology* 9:397-423.

Kasinitz, Philip. 1992. *Caribbean New York: Black Immigrants and the Politics of Race*. Cornell University Press.

Korenman, Sanders, and David Neumark. 1991. "Does Marriage Really Make Men More Productive?" *Journal of Human Resources* 26:282-307.

Lieberson, Stanley, and Lawrence Santi. 1985. "The Use of Nativity Data to Estimate Characteristics and Patterns." *Social Science Research* 14:31-56.

Mincer, Jacob. 1974. *Schooling, Experience, and Earnings*. Columbia University Press.

Model, Suzanne. 1991. "Caribbean Immigrants: A Black Success Story?" *International Migration Review* 25:248-76.

Palmer, Ransford W. 1974. "A Decade of West Indian Migration to the United States, 1962-1972: An Economic Analysis." *Social and Economic Studies* 23:571-87.

Reid, Ira De A. 1939. *The Negro Immigrant: His Background, Characteristics and Social Adjustment, 1899-1937*. Columbia University Press.

Sowell, Thomas. 1978. "Three Black Histories." Pp. 7-64 in *Essays and Data on American Ethnic Groups*, edited by Thomas Sowell with Lynn D. Collins. The Urban Institute.

Stevens, Gillian, and Joo Hyun Cho. 1985. "Socioeconomic Indexes and the New 1980 Census Occupational Classification Scheme." *Social Science Research* 14:142-68.

Tidrick, Kathryn. 1971. "Need for Achievement, Social Class, and Intention to Emigrate in Jamaican Students." *Social and Economic Studies* 20:52-60.

Tienda, Marta, and D. Lii. 1987. "Minority Concentration and Earnings Inequality: Blacks, Hispanics, and Asians Compared." *American Journal of Sociology* 93:141-65.

U.S. Bureau of the Census. 1992. *Public Use Microdata Samples: Technical Documentation*. Government Printing Office.

Waters, Mary C. 1991. "The Role of Lineage in Identity Formation among Black Americans." *Qualitative Sociology* 14:57-76.

Wilson, William Julius. 1987. *The Truly Disadvantaged*. University of Chicago Press.

Deprivation and Race Riots: An Extension of Spilerman's Analysis*

SUSAN OLZAK, *Stanford University*
SUZANNE SHANAHAN, *University College Dublin*

Abstract

Many scholars currently link the growth of a black underclass to the persistence of racial tensions and riots in American cities. However, nearly all previous systematic studies of race riots refute these claims. We propose an alternative explanation that rests on competition processes. In this article we extend Spilerman's 1971 analysis of community characteristics and race riots to include 1954-93. Our results parallel earlier ones that found little evidence that black poverty or racial disparities in income, education, and occupation caused race riots. Furthermore, we find that the effects of black poverty are generally opposite from what most deprivation theories would lead us to expect. Both low income for blacks and dilapidated housing significantly depress the rate of racial unrest. Finally, we find that competition that is expressed as the contraction of job opportunities for minorities created by demographic pressures and unemployment significantly increases the rate of race riots.

For decades social scientists were convinced that disparities between races in economic opportunity generated racial unrest (Blauner 1971; Chicago Commission on Race Relations 1922; Downes 1969; Grimshaw 1960; Lee & Humphrey 1943; Morgan & Clark 1973; Myrdal 1944; U.S. National Advisory Commission on Civil Disorders 1968). The majority of systematic studies on race riots during the 1960s found little support for a causal connection. Regardless of the more recent findings, such ideas continue to influence public perceptions of the causes of riots (e.g., *The Economist* 1992).

** Preparation of this article was supported by a grant to the first author from the National Science Foundation (SES-9196229). Expert research assistance and suggestions were provided by members of the Ethnic Collective Action Workshop at Stanford University: Andy Almazan, Lisa Brooke, Mateo Cordova, Sergio Garza, Angela Jones, Amy Lanigan, Ruth Lopez, Adam Miller, and Barry McMillon. We would like to extend our appreciation to Seymour Spilerman, who graciously sent us his original data on community characteristics. Finally, colleagues Michael Hannan, Elizabeth H. McEneaney, John Meyer, Daniel Myers, and William Tobin offered extremely useful criticisms and suggestions on previous drafts of this article.* *Direct all correspondence to Susan Olzak, Department of Sociology, McClatchy Hall, Stanford University, Stanford, California 94305.*

 Social Forces, March 1996, 74(3):931-961

It is intriguing that explanations of unrest by means of poverty have survived despite the overwhelming number of studies that contradict them (e.g., Bloombaum 1968; Lieberson & Silverman 1965; Spilerman 1970b, 1971, 1976). Although black poverty is widespread in urban settings, deprivation levels apparently do not distinguish between riot and nonriot cities (e.g., Bloombaum 1968; Spilerman 1970a). Nor does a widening gap in income or occupational prestige predict riot severity (Spilerman 1976). Nevertheless, grievance and poverty models of unrest became the standard explanation for race riots located in black ghettos (Boskin 1976; Feagin & Hahn 1973).

Recent outbreaks of racial violence (e.g., in south central Los Angeles or Liberty City in Miami) have stirred renewed interest in deprivation explanations. Because these race riots occurred in minority communities with high rates of unemployment, widespread poverty, and poor housing, deprivation explanations seem pertinent to many observers. To date, no cohesive argument has been offered that is consistent with both prior findings and current examples of racial unrest.

In this article we examine an alternative to deprivation theory that is based upon collective action perspectives. We examine the effects of both competition processes and the changing urban conditions of poverty, occupational segregation, and the shrinking stock of housing for low-income minorities. We also offer analysis of new data on race riots in cities from 1954 to 1993, so we can assess the generalizability of previous findings. Our research analyzes the rate of race riots across 204 of the cities with the largest populations in four decades of American race relations.

The empirical analysis that follows highlights one implication of the competition perspective of ethnic collective action — that changes in the size of the minority population and economic contraction together spark ethnic/racial conflicts. To address this point, our analysis uses information on urban poverty, city political structure, and economic indicators in models of racial unrest in cities. In conducting this research we were fortunate to have access to Spilerman's original data on poverty and political characteristics from 1950 to 1960 at the city level. In this article we combine Spilerman's information on urban poverty, population composition and change, and political structure with original data collected on racial unrest.[1]

Defining Ethnic/ Racial Collective Action

Previous research on ethnic collective action has distinguished between two broad categories of events: race/ethnic protests and race and ethnic conflicts (Olzak 1992). Ethnic protests are public, collective events in which an ethnic group makes a claim or articulates a grievance on the basis of ethnic or national identity. The claim is specifically ethnic and is made on behalf of the ethnic group as a whole. Protests are usually on behalf of one group, which is making a request to change racial policies, or to register a grievance, as in claims of police brutality. Most protests are also distinguished by the fact that violence may occur but it is usually not instigated by the protestors themselves. The civil-rights protest is a modern prototype of an ethnic protest.

The second broad category is ethnic conflicts, in which two or more ethnic groups are engaged in a collective confrontation with each other, or an ethnic group attacks a specific group as a symbol of authority or law, such as the police or National Guard. Activity usually involves an attack on group members (or an institution symbolizing that group, such as a synagogue) based solely on the ethnic or racial identity of the group. Graffiti or epithets or insults will be used to denigrate another group's language, skin color, or nationality. Conflicts may, but do not necessarily, involve violence or the threat of violence to people or property. Ethnic conflicts are distinguished from protests by the presence (real or symbolic) of an ethnic target of the event and the presence of a threat or actual use of violence to attack a group. Using these definitions of ethnic collective action, it becomes clear that large-scale riots could emerge during the course of either protests or conflicts.

The possibility that protests or conflicts could lead to rioting prompts a new question: How are race riots different from racial protests and conflicts? The answer to this question is not straightforward. This is partly because race/ethnic violence has taken many forms in many settings, in many countries and partly because riots can involve more than one race or ethnic population.[2] Another complication documented in previous research on racial unrest is that racial violence instigated by different race/ethnic groups reveals very different causal patterns (Olzak 1992; Tolnay & Beck 1995). Finally, there are important considerations of maximizing comparability across studies. For example, Lieberson and Silverman (1965), Spilerman (1970b, 1971, 1976), and Wanderer (1969) analyzed instances only of black race riots. We find this last issue to be a compelling one and so we follow these researchers and restrict our definition of race riots to spontaneous, large-scale, and public displays of racial violence where there was property damage, injuries, or deaths. Further, to ensure comparability with earlier research on American race riots, we restrict our analysis to events in which blacks initiated acts of aggression against property, persons, or symbols (Spilerman 1970a:630).

Collective Action Perspectives and Race Riots

ABSOLUTE AND RELATIVE DEPRIVATION PERSPECTIVES

Two branches of deprivation theory have informed the study of race riots. According to absolute deprivation formulations, cities where high proportions of blacks have low levels of education or are employed in highly segregated and low-paying jobs are more likely to experience racial unrest (Chicago Commission on Race Relations 1922; Park 1949; U.S. National Advisory Commission on Civil Disorder 1968). According to this view, the dramatic rise of race riots in Newark, Chicago, Detroit, and other cities during the mid-1960s seemed to provide strong confirmation of these theories.

Deprivation theories of racial unrest have reemerged in the literature on the urban underclass and unrest. Recent analyses of the construction of black ghettos suggest that black poverty has become widespread at the same time that racial tensions remain high (Harris & Wilkins 1988; Massey & Denton 1993). In this view, the concentration of poor families in communities where African

Americans and other poor minorities have been segregated encourages racial grievances that sporadically erupt into episodes of fire bombing, looting, or armed violence. For example, Massey and Denton (1993), observe

> The economic deprivation, social isolation, and psychological alienation produced by decades of segregation bore bitter fruit in a series of violent urban riots during the 1960s. The violence began in Birmingham, Alabama, in the summer of 1963, but the real bellwether was the Los Angeles riot of August 1965, which did $35 million worth of damage and left 4,000 injured and 34 dead. After sporadic violence in Chicago and Cleveland during the summer of 1966, a convulsive wave of mob violence erupted during July and August of 1967, when black ghettos in sixty U.S. cities exploded in a cataclysm of frustration and rage. (58)

Eisinger (1973), Graham and Gurr (1969), Gurr (1972), and others offer a revised version of deprivation theory to explain why racial unrest occurs during periods in which minorities are becoming less deprived. This related line of argument emphasizes that relative disparities in education, income, and occupations for whites and blacks ignite unrest. One popular version of the relative deprivation argument holds that racial unrest arises when the black/white gap in attainment grows at the same time that minority expectations are rising. Political scientists such as Gurr and Eisinger emphasize that grievances are fundamentally political, driven by the lack of political power and absence of local control over city government and services in minority communities. In contrast, sociologists tend to emphasize poverty and unemployment as key structural features of racial discrimination leading to unrest (e.g., see Coser 1972).

LIMITATIONS OF DEPRIVATION THEORIES

Advocates of deprivation theories have met criticism from several sources who contend that deprivation theories have failed to offer a mechanism linking deprivation, grievances, and the inclination to participate in riots (Keily & Isaac 1984). Others are doubtful because race riots peaked during a period in which average levels of income (by educational levels) for blacks were approaching those of whites (Wilson 1978). The trajectories of these two temporal processes are inconsistent with deprivation theory expectations.

In general, deprivation theories have not fared well in most systematic empirical investigations of the topic. In particular, few researchers find evidence that underclass poverty was associated with participation in riots (Paige 1971; Spilerman 1970a, 1971). Researchers find that cities with less black poverty experience higher rates of race riots than do cities with higher proportions of blacks in poverty. Concentration of poverty in minority populations does not appear to lead to protest against such conditions (Spilerman 1970a, 1971, 1976). Moreover, it appears that integration in housing and occupations increases racial violence under some conditions (Grimshaw 1969; Wilson 1987).

Deprivation explanations of racial unrest in general are often limited by a reliance on case studies. Deprivation explanations evoke powerful images that can describe how racial injustice accumulates to reach intolerable levels, but attributing cause to descriptions of conditions in riot cities is problematic because such descriptions ignore what conditions may hold in cities without riots (Spilerman 1970b). Most case studies that emphasize racial poverty suffer

from selection bias in that they are based on cities that experienced riots (e.g., Ladner et al. 1981). This type of analysis describes the preconditions in a riot city but cannot explain the process.

Other researchers investigating causal factors have surveyed riot participants after outbreaks of racial violence. (See Oliver & Johnson 1984; Paige 1971.) Establishing a causal link between reports of grievances and outbreaks of violence *after* a riot has proven difficult. One could easily imagine the reverse causal sequence, in which rioting sharpens the focus of poverty conditions into political grievances (Piven & Cloward 1977).

Although deprivation theories are flawed, they continue to affect public policy, academic debate, and public opinion on the causes of race riots. More than twenty years have elapsed since most of the key studies of race riots at the local level were published.[3] So it seems to be an opportune time to reinvestigate deprivation explanations of racial unrest. Since these studies were published, new developments in collective action theory and methods have proved useful to the study of collective unrest. (See McAdam, McCarthy & Zald 1988 and Olzak 1989 for reviews.) In the following section we apply some of these perspectives to examine deprivation arguments through the lens of collective action theory to provide some new insights into the causes of racial unrest in the contemporary period.

RESOURCE MOBILIZATION THEORY AND RACIAL UNREST

Several variants of resource mobilization theory emerged partly to fill the gap created by the empirical failure of deprivation perspectives (Tilly 1978). In particular, resource mobilization and competition theories suggest that declining (rather than growing) inequality facilitates higher rates of mobilization (Blalock 1967; Jenkins 1983; Tilly 1978). The resource mobilization perspective draws attention to processes of social change that redirect resources away from powerful groups to more disadvantaged ones. By this argument, under conditions of rising resources and declining disadvantage, solidarity and mobilization by the powerless or challenger groups is encouraged.

Using resource mobilization perspectives to analyze social protest has been useful in identifying forces of change that facilitate collective action. The prevailing view from leading theories of collective action would lead one to expect that protest activity in particular occurs during periods of economic expansion rather than deprivation. If we follow Baldassare (1994), Harris and Wilkins (1988), Rossi (1973), and others who consider race riots broadly as a form of racial protest, then theories of racial protest become relevant to this study of racial unrest. In particular, Morris (1984) successfully applied a version of resource and organizational theories to racial protest. He demonstrated that indigenous resources within the black community (notably student and religious organizations) that were mobilized by African Americans became channeled into a variety of protest movement organizations.

Recent reviews of collective action theory underscore the importance of many of the insights drawn from resource mobilization theorists. Resource mobilization theorists focus on the dynamics of organizational growth and decline in social movements and movement activity (McCarthy, McAdam &

Zald 1988). Such an emphasis has been a welcome shift away from static models and conditions of poverty or emotional states of deprivation.

Although they are extremely informative, resource mobilization theories do not directly address the question of why improving conditions would produce racial unrest rather than mobilization along class, occupation, gender, or some other boundary. To answer this question, we combine aspects of resource mobilization theories with competition perspectives of race relations to explain the conditions under which minorities or disadvantaged groups mobilize racial protest, conflict, and rioting.

COMPETITION PERSPECTIVES AND RACIAL UNREST

Until recently competition theories have focussed on instances of racial and ethnic unrest in which the advantaged groups attack minorities to contain or restrict competition (Blalock 1957, 1967; Bonacich 1972; Olzak 1992; Spilerman 1970a). This article shifts the focus of competition theory to consider the question of whether competition fuels collective action by the disadvantaged as well.

Competition theories of race relations build on Barth's (1969) insight that displacement and exclusion are generated by competition. It assumes that initial overlap of habitats intensifies competition among groups, which in turn encourages attempts to exclude the competitors. Such attempts at exclusion often entail violent conflict.[4] Competition theories treat race or ethnicity as a *social boundary*, commonly based upon characteristics such as geographical origin, race, religion, language, or some other cultural symbol. Which (if any) racial or ethnic markers become emphasized depends upon the context of interaction (Barth 1969). Race riots are considered here as an extreme form of ethnic collective action, in which large-scale violence and public confrontations based upon racial grievances become widespread.

The ecological tradition focuses on competition as a key social process that activates and deactivates ethnic identities and boundaries. Competition shapes collective action organized around ethnic boundaries, and ethnic collective action in turn affects the persistence and salience of ethnic boundaries (Olzak & West 1991).[5] We follow Blalock (1967), who defines ethnic competition as

> involving the idea that two or more individuals are striving for the same scarce objectives, so that the success of one implies a reduced probability that others will also attain their goal. (73)

One important implication of this view is that competitors do not need to be aware of each other's existence. Nor does the tradition assume that competition inevitably produces conflict and strife. Instead, the theories specify that the rate of ethnic conflict depends on the degree of niche overlap. Barth (1969) develops this point in distinguishing the conditions that determine stable versus unstable ethnic relations:

> Stable inter-ethnic relations presuppose such a structuring of interaction: a set of prescriptions governing situations of contact. . . . [but] when two or more interspersed groups are in fact in at least partial competition within the same niche . . . one would

expect one such group to displace the other, or an accommodation involving an increasing complementarity and interdependence to develop. (16-20)

Competition theorists maintain that the likelihood of racial and ethnic conflict is low when ethnic and racial groups occupy nonoverlapping (or segregated) regions of social structure (Barth 1969; Olzak 1992). An increase in niche overlap has important consequences for conflict: when groups come to occupy the same niche, the historically more powerful or advantaged group attempts to exclude competitors. When the less powerful resist these attempts, racial conflict and violence ensues.

Competition perspectives add that forces that break down racial barriers also encourage racial competition among formerly segregated groups. If breakdown of racial segregation is accompanied by decreasing racial inequality, then mobilization by disadvantaged groups is likely to occur. Thus both resource mobilization and competition theories of ethnic mobilization hold that competition forces could have two consequences: It can generate waves of unruly protest (on the part of disadvantaged minorities) in the form of race riots or rising competitive forces can cause reactive mobilization on the part of dominant groups. (And the two can be mutually reinforcing, as in movements and countermovements.)

What kinds of changes would produce more racial competition? Economists and economic historians generally argue that tight markets imply that overall labor market competition is intensifying when there are signs that the economy is becoming weak (compared to periods of growth or recovery — Higgs 1977). For race/ethnic groups, a rise in minority unemployment rates would signal a downturn in their fortunes (Jencks & Peterson 1991). To the extent that recessions and depressions increase pressure to find jobs that are less plentiful, competition among minorities in the same job niche increases. Economic contraction may also force those with middle-level jobs to search for unskilled jobs, thus pitting older, upwardly mobile cohorts against each other — or perhaps competition with earlier waves of immigrants or migrants increases (Wilson 1978). Factors that indicate a downturn in the economy are therefore expected to produce ethnic collective action in response to a shrinking job market. By this argument, we would expect that unemployment generates higher levels of collective violence instigated by minorities.

Competition theories, including split labor market formulations, further specify the relationship by emphasizing that it is not job segregation and economic hardship per se that activates ethnic antagonism (Bonacich 1972). Rather, ethnic antagonism is due to economic factors that raise the supply of low-wage labor. This in turn increases competition among ethnic groups (Blalock 1957, 1967). In retaliation, groups affected by hardship direct attacks against the group believed to be the source of that competition. To the extent that the disadvantaged resist becoming victimized, racial unrest escalates.

One implication of the split labor market formulation is that the combination of a rising supply of ethnic/racial labor willing to work at lower than prevailing wages and economic contraction intensifies ethnic competition.[6] That is, these formulations imply an interaction effect between a declining set of job opportunities and immigration affecting the rate of ethnic collective action. Thus

competition theories contrast with economic hardship formulations in arguing that economic decline does not inevitably aggravate ethnic relations. Rather, they argue that unemployment, in combination with high proportions of minorities, raises levels of ethnic competition, which in turn increases rates of ethnic collective action, at least initially.

Our overall argument holds that competition elevates ethnic tensions under two similar conditions of rising competition. The first condition would be one of general elevation of competition in the labor market as a whole: when population increases at the same time that unemployment rises, ethnic competition rises overall. In the second, economic contraction within the minority niche would reduce the number of jobs open to immigrants, minorities, and other less-skilled workers, raising the level of competition for fewer openings in low-paying jobs. To the extent that the minority population in particular rises, competition involving minorities would be intensified when these conditions hold because demands for jobs and housing in minority communities would be increasing at the same time that resources for this group would be shrinking.

Thus the main goal of this article is to evaluate the usefulness of a competition explanation of racial unrest. In doing so we attempt to reconcile the assumptions of poverty and deprivation theories with past research findings.

Research Design and Measurement

Our research analyzes the count of race riots that occurred in any one of 204 largest population cities in the U.S. from 1954 to 1993. The 204 cities represent all cities for which full data on events, poverty, population composition, and political structure exist.[7]

TIME FRAME

We add to Spilerman's research on race riots (which reported analysis of the 1961-68 period) by beginning our observations in 1954 and then extending our analysis into the 1990s. While only a few riots occurred before 1961, we begin our observations on racial unrest in 1954, which marks an important institutional shift in American race relations: it marks the year of the landmark Supreme Court decision, *Brown* v. *Board of Education of Topeka,* mandating that all public schools begin to desegregate "with all deliberate speed." Our data include initial reactions to this ruling, some of the first civil-rights movement activities in the early 1950s, racial unrest in urban centers during the 1960s, and other race/ethnic confrontations (including the 1992 Los Angeles race riot).

NEWSPAPER SOURCES

Our research design tracks histories of all instances of civil disorders that met our criteria and were reported in the *New York Times.*[8] We first identified "candidate events" from the NYT Index and then coded characteristics of race riots from daily microfilm accounts. Our coding scheme judges that a race riot occurred if the event involved racial grievances against discrimination or perceived racial injustice and in which 30 or more persons engaged in violent

activity that lasted several hours. In our coding scheme, the distinguishing feature of race riots was the articulation of racial grievances voiced by the participants, but this is not the sole characteristic of race riots. In most (but not all) cases, substantial violence occurs, often involving some combination of injuries, looting, arson, sniping, and damage to property.[9] In all cases, the police were present in large numbers. We reasoned that because racial deprivation theories assume that gaps between the economic conditions of races drive unrest, it made sense to focus on poverty characteristics and race riots within the same community. Here we analyze race riots in which African Americans were the main participants.

Daily reports from the *New York Times (NYT)* are collected in two stages. We first identified "candidate events" from the *NYT* Index, determined whether they fell under our coding criterion, and then coded characteristics of events from microfilms of daily accounts. Following an established tradition in collective action research, the research project used newspaper accounts coded from daily microfilms of the *New York Times* to provide continuous coverage of events in cities.[10] Several empirical investigations of possible sources of bias have shown convincingly that national newspaper accounts have less reporting bias than local accounts, contrary to original suspicions (Snyder & Kelly 1977). On the other hand, systematic comparisons of newspaper accounts to official statistics (using records of police permits) show that newspapers are likely to be influenced by the size of protests (McCarthy, McPhail & Smith n.d.).

More recently, Olzak and Brooke (1995) compared more than twenty riot accounts in the *New York Times* to those found in several local newspapers and in one aggregated newspaper data bank.[11] Because of findings in earlier studies, these authors thought it prudent to consider the possibility that local sources differ from *New York Times* accounts in systematic ways. They compared accounts of events (or nonevents) in Detroit, Chicago, Miami, and Atlanta in local sources with accounts in the *New York Times*. They also compared two local black newspapers with local mainstream presses. Their comparison of local, regional, and *New York Times* accounts of race riots found surprisingly few discrepancies that matter to event-history analysis. No major differences among the *Atlanta Constitution, Miami Herald,* and the *New York Times* were found for key variables of the timing, duration, number, and ethnicity of those involved or the location of riots. There were some discrepancies in arrest records and estimates of police presence, but these did not vary by more than 10% (which can be attributed to the timing of newspaper deadlines in combination with time zone differences).

INDEPENDENT VARIABLES

We analyze the effects of all independent variables at the city level provided by Spilerman in his original analysis of racial disturbances (Spilerman 1970a). The political structure measures and black occupation, education, and income variables are from Aiken and Alford (1970). A complete set of these variables is available for 204 U.S. cities. All indicators are from the 1960 census with the exception of dilapidated housing, which is available only for 1950, the change

in nonwhite population from 1950 to 1960, and the change in total population from 1950 to 1960.[12]

INDICATORS OF SOCIAL DISORGANIZATION

As noted in Spilerman (1970a, 1971), social disorganization arguments hold that in-migration of minorities and rapid population growth induce strain on resources and ethnic tensions among new residents. Furthermore, a higher percentage of unlivable housing units puts a strain on city resources, constituting a major grievance for the poorest minority residents (Wilson 1987). According to the disorganization perspective, all three indicators would produce higher levels of racial unrest.

Three measures of community disorganization include the change in total population from 1950 to 1960, the percentage change in the nonwhite population during the same period, and the percentage of dilapidated housing units. Dilapidated housing is defined as the percentage of boarded-up housing units in a city (U.S. Bureau of the Census 1950).

INDICATORS OF DEPRIVATION

As discussed above, both absolute and relative deprivation arguments have been offered as key causes of racial turmoil. Four indicators of absolute deprivation in the nonwhite community are percentage of nonwhites in low-status occupations (household workers, service workers, and common laborers), median years of education for nonwhites, median family income for nonwhites, and nonwhite unemployment rate. Deprivation arguments would hold that nonwhite income and nonwhite education measures depress the rate of race riots while nonwhite participation in low-status occupations and nonwhite unemployment would raise the level of racial unrest.

As Spilerman (1970a) argued, one variant of these theories suggests that poverty produces grievances and unrest to the extent that it affects the minority community disproportionately. In this view, poverty that falls equally on whites and nonwhites might produce class mobilization rather than ethnic and racial tensions. If relative deprivation views are correct, racial unrest would be more likely where the black poverty rate exceeds the white poverty rate. Relative deprivation arguments suggest that racial unrest erupts where there is a substantial racial gap in income, education, and occupation. Four variables are used to test relative deprivation. Each measures the relative position of nonwhites vis-à-vis whites in a community: (1) ratio of the percentage of nonwhite males working in traditionally black occupations to the percentage of white male laborers employed in similar occupations, (2) the ratio of nonwhite median family income to white median family income, (3) the ratio of nonwhite male unemployment rate to the white male unemployment rate, and (4) the ratio of nonwhite median education to white median education. Following relative deprivation arguments, the ratio of black/white income and education should vary inversely with the likelihood of race riots; that is, race riots will decline as the racial gap in attainment is narrowing. In this same view, a higher concentration of nonwhites in low-status occupations and higher participation

in unemployment would generate more racial violence in those communities most affected by these gaps in economic well-being.

MEASURES OF CITY POLITICAL STRUCTURE

In addition to stressing the economic conditions in minority ghettos, Lieberson and Silverman (1965), the U.S. National Advisory Commission (1968), and others emphasized preconditions that structurally blocked access by the black community to routes of political power. The National Advisory Commission (1968) mentioned a combination of at-large elections and "weak mayor" systems as less accountable to minority constituents. Furthermore, the commission concluded that city governments were less sensitive to minority communities when the ratio of city council members to constituents was low. To capture the effects of political sensitivity to minority grievances and concerns, Spilerman (1970a) included four measures of city political structure: population per councilman; percentage of councilmen elected at large; a dummy variable of partisan versus nonpartisan election; and a dummy variable for mayor-council form of city government form of government versus any other form. The general argument tested by Spilerman was that cities that were less responsive to the concerns of minorities would experience more unrest. These characteristics imply that when cities have a lower population per council member and partisan elections, then politicians are likely to be sensitive to racially homogeneous neighborhoods and organizational constituencies (see also Aiken & Alford 1970).

Many arguments about the effects of political structure on race riots rest on the issue of government accountability. For example, Spilerman (1970a) argues that the mayor-council governments would be more responsive to minority concerns, because they could be held accountable by the minority communities directly through elections than would appointed city managers or other forms, such as at-large elections, that could dilute minority influence. Following Spilerman, we expect at-large elections, the ratio of population to council members, and the presence of nonpartisan elections to increase the likelihood of race riots, but the presence of a mayor-council form of city government to decrease the chances of racial unrest.

INDICATORS OF RACIAL COMPETITION

Our competition arguments imply that contraction of the economy produces higher rates of ethnic/racial conflict, as groups vie for a shrinking number of jobs and resources. In addition, resource mobilization and competition arguments suggest that a decline in racial disparities in income ought to produce higher rates of conflict. As formerly disadvantaged minorities come to compete in new job and housing markets, advantaged majority groups attempt to exclude the competitors.

To examine our arguments empirically, we include two measures of competition. The first competition measure is specified as an interaction term between percentage change in population and the overall unemployment rate. This is a measure of *overall competition* in the labor market. The second measure

is *minority competition*. It is specified as an interaction between the nonwhite unemployment rate and percentage change in the nonwhite population in a city from 1950 to 1960. We include both the main effects and interaction terms in our investigation of competition.

In contrast to pure deprivation arguments, theories emphasizing both resource mobilization and competition would hold that as nonwhite income rises relative to white family income, racial protest of all kinds should increase. To the extent that this resource competition argument also holds, we would expect that the rate of race riots would also peak as a result of declining income disparity between the races.[13]

DEPENDENT VARIABLE: EVENT COUNTS

The dependent variable in our analysis uses the distribution of riot counts as grouped events occurring on a particular date to capture the intensity of race/ethnic conflict (Aalen 1978; Petersen 1991). That is, the observed count is aggregated to city counts. In this sample of 204 cities, the mean number of riots was 1.22. Fifty-eight percent (118 cities) had no race riots, which means that 86 cities had at least one race riot over this period. Only two cities, New York and Chicago had a substantially large number of race riots, tallying 23 and 19, respectively. Furthermore, our definition of race riots as large-scale, violent events holds for the observed set of race riots. Virtually all race riots involved fifty or more participants (and 17% had more than 500 participants). In all but three riots, police or bystanders sustained injuries, or there was substantial damage to buildings, or participants were reported as using bricks, stones, firebombs, or other weapons. Targets of race riots were varied, but nearly one-fourth (23%) of all riots observed had the system of justice, city government, or the police as targets of rioting behavior. Another 60% of the observed riots had no clear-cut target or symbol that could be identified from newspaper accounts. The remaining race riots involved a grievance against one particular ethnic group identified as the reason for the rioting (e.g., whites, Jews, or Mexican-Americans).

To capture the environmental causes of variation in outbreaks of racial unrest, we follow Spilerman's lead and estimate effects of covariates on the rate within the framework of Poisson regression.[14] That is, we form regression models for the (daily) count of ethnic events, denoted here by y_t. The regression model represents the rate, λ_t, as a function of observed covariates. To ensure that λ_t can only take nonnegative values, it is parameterized as a loglinear function of the covariates:

$$\lambda_t = \exp(x'_t\beta) \tag{1}$$

This regression model assumes that each observed count of race riots is a realization of a Poisson process whose rate is given by equation 1. In other words, the conditional probability function at time t is given by:

$$Pr(Y_t = y_t | \lambda_t) = \frac{\lambda_t^{y_t} e^{-\lambda_t}}{y_t!} \tag{2}$$

where λ_t is the mean and the variance of the distribution of the dependent variable, i.e.,

$$E(Y_t) = \mathrm{Var}(Y_t) = \lambda_t \tag{3}$$

The Poisson assumption that the expected value of Y_t is equal to its variance is often not justified (Barron 1992; Cameron & Trivedi 1986). In order to test this assumption we also estimated the rate of collective action using negative binomial regression, which allows for the possibility that the variance of the dependent variable is greater than its mean. However, these models did not fit significantly better than the Poisson models, so we report only the Poisson models with heteroskedastic-consistent errors.

Results

We first compare bivariate relationships between covariates and riots counts in our analysis with Spilerman's (1971) results. We then turn to our analysis of the effects of poverty followed by our analysis of labor force competition and racial unrest in the final table.

BIVARIATE CORRELATIONS IN TABLE 1

Virtually all the measures of bivariate association between deprivation and riots shown in Table 1 indicate a lack of support for deprivation arguments. It is striking how similar in magnitude and direction the coefficients are across all three columns. We found that in no case were the signs opposite and, in most cases, the magnitude of the correlations were similar. As Spilerman found, population size, size of the nonwhite population, and southern location are significantly correlated with event counts, while nearly all indicators of social disorganization are negatively correlated with racial unrest. Reflecting the same pattern as in Spilerman's results, we see that while percentage of nonwhite males in low-status occupations is significantly and negatively correlated with race riots, higher median income in both studies were significantly and positively correlated with the count of race riots.

The indicators of political structure show that riot counts are slightly more correlated with political indicators in our riot data than in Spilerman's correlation coefficients in column 1. In particular, we found that the percentage of city-council members elected at large was negatively and significantly correlated with the riot count (where Spilerman reports a nonsignificant but negative correlation with riot counts). However, all the signs of the coefficients remain the same, with the population per councilmen (as a measure of unresponsive government) significantly correlated with race riots, as was the presence of a mayor-council government. Taken together, the pattern of correlations seems remarkably similar across the two studies. Such parallel bivariate relationships lend credibility to the comparisons that follow.

Given the literature on the "urban underclass," it might seem surprising that nonwhite population size and black poverty are not more highly correlated with each other. It is perhaps useful to recall that the poverty measures are taken in 1959, long before the poverty programs were introduced that sharply

TABLE 1: Bivariate Correlations of Race Riot Counts and Community Characteristics

	Spilerman 1970 (1)	1961-1968 Riots (2)	1954-1992 Riots (3)
Region and population size			
South (0,1)	-.198*	-.178**	-.148+
Northeast (0,1)	na	.132	.191**
North central (0,1)	na	.097	.030
West (0,1)	na	-.019	-.042
Ln Percent nonwhite population	.843**	.170**	.227**
Ln Population	.586*	.515**	.581**
Indicators of social disorganization			
Percent change in total population 1950-1960	-.144*	.124+	-.135+
Percent change in nonwhite population 1950-1960	.044*	-.010	-.012
Percent dilapidated housing 1950	-.180**	-.160*	-.171**
Indicators of absolute deprivation			
Percent nonwhite males in low-status occupations	-.215*	-.214**	-.214**
Percent nonwhite males unemployed	.067	.062	.054
Nonwhite median-family income in 1,000s	.182*	.204**	.183**
Nonwhite median education	.078*	.038	.032

+ $p < .10$ * $p < .05$ ** $p < .01$

TABLE 1: Bivariate Correlations of Race Riot Counts and Community Characteristics (Continued)

	Spilerman 1970 (1)	1961-1968 Riots (2)	1954-1992 Riots (3)
Indicators of relative deprivation			
Percent nonwhites males in low-status occupations divided by white figure	-.166*	-.196**	-.256**
Nonwhite median family income divided by white figure	.148*	.134+	.142*
Nonwhite unemployment rate divided by white figure	-.032*	-.024	-.020
Nonwhite median education divided by white figure	.151*	-.160+	.162*
Indicators of political structure			
Population per council members	.485*	.481**	.562*
Percentage of city council elected at large	-.069	-.207**	-.174*
Nonpartisan elections (0,1)	-.016	-.006	-.032
Presence of mayor-council government (0,1)	.184*	.189*	.192**
Indicators of competition			
Interaction of increase in population and total male unemployment 1950-1960	na	-.095	-.095
Interaction of increase in black population and black unemployment 1950-1960	na	.021	.019

+ $p < .10$ * $p < .05$ ** $p < .01$ (two-tailed tests)

reduced poverty in the South and Southwest.[15] Thus, as Wilson (1987) reminds us, poverty in 1959 was more rural and less racially concentrated in urban ghettos than in the 1990s.[16]

NONWHITE POPULATION, POVERTY, AND POLITICAL STRUCTURE

Did racial deprivation raise the rate of race riots over the 1954-93 period? Table 2 shows little evidence that it did.[17] Neither low levels of attainment nor relative gaps in attainment for blacks apparently affect the likelihood of race riots.[18]

Furthermore, the effects of percentage of dilapidated housing and gap in black/white income are the reverse of what deprivation arguments would hypothesize. That is, cities with less dilapidated housing and cities where nonwhites have incomes closer to whites experience higher rates of race riots. The results suggest that improving rather than declining conditions in the black community are associated with race riots. Thus the results in Table 2 are consistent with resource mobilization and competition theory arguments and inconsistent with most deprivation theory arguments.

A notable exception to this statement is the effect of black unemployment. We find in columns 4 and 5 that black unemployment has significant and positive effects. That is, in two out of five models in Table 2, cities with higher rates of unemployment for blacks had significantly higher rates of unrest. This finding suggests that male unemployment plays a role in producing racial tensions. The fact that black unemployment has a significant effect on race riots is also consistent with arguments that black unemployment is a major determinant of the growth of the black underclass (Wilson 1987).

The remainder of our results that show no effects of racial poverty are similar to those reported over twenty years ago by Feagin and Hahn (1973), Lieberson and Silverman (1965), and Spilerman (1971). Extremes in racial poverty or in gaps in attainment do not increase the rate of race riots. Quite the opposite pattern appears to hold.[19]

Other factors also show effects parallel to those found decades ago. City size and the size of the black population have enormous and significant effects on the magnitude of racial unrest in all five columns. These effects are potent and consistent throughout our analysis, even when regional and poverty effects are included in the model.[20] The coefficient of .595 tells us that when the black population is one standard deviation higher than the mean, the rate of race riots increases by 65%. For this reason we include city size and the size of the nonwhite population as controls in all our models.[21]

Two other key indicators of regional and political differences that have powerful effects in Table 2 were unrelated to race riots in Spilerman's analysis. While Spilerman found a negative and significant effect of southern location, we did not. Instead, our results show powerful and positive effects of northeastern location, even when other population and poverty variables are included in the model. In addition, we find that nonpartisan elections produce higher rates of race riots, suggesting support for the original structural responsiveness argument. In this view, partisan electoral politics would be more responsive to minorities who would provide distinct constituencies that would hold local

TABLE 2: Poisson Regression Effects of Racial Deprivation on Race Riots, 1954-1993

	(1)	(2)	(3)	(4)	(5)
Intercept	-8.34**	-6.74**	-8.78**	8.21**	-8.36**
	(.709)	(.877)	(.737)	(.777)	(.715)
Ln Population 1960	.712**	.643**	.663**	.707**	.704**
	(.066)	(.053)	(.051)	(.050)	(.057)
Ln Percent nonwhite population 1960	.595**	.692**	.705**	.557**	.604**
	(.105)	(.110)	(.111)	(.142)	(.111)
Percent change in nonwhite population 1950-1960	.228	.161	.199	.223	.227+
	(.193)	(.116)	(.113)	(.117)	(.116)
Percent nonwhite males unemployed 1960	.223	.161	1.02	2.22*	2.08*
	(.116)	(.119)	(1.88)	(1.77)	(1.85)
Nonpartisan elections	.513**	.494+	.464**	.505**	.507**
	(.176)	(.174)	(.177)	(.177)	(.177)
Northeast (0,1)	.758**	.616*	.612*	.738**	.737**
	(.326)	(.186)	(.185)	(.190)	(.199)
Percent dilapidated housing 1950		-4.78*			
		(1.63)			
Ratio of nonwhite/white median income 1960			2.33**		
			(.799)		
Ratio of nonwhites/whites in low-status occupation				-.067	
				(.172)	
Ratio of nonwhite/white median years education					.207
					(.767)
Log likelihood	-42.04	-37.54	-37.80	-41.94	-41.98

[a] Heteroskedastic-consistent standard errors are in parentheses.

+ p < .10 * p < .05 ** p < .01 (two-tailed tests)

government accountable for their actions. The effects do support the view that nonpartisan elections raise the rate of racial unrest.

The positive effects found for northeastern location are equally consistent with another ad hoc explanation offered by urban historians. The consistently significant and positive effects found for northeastern location (in all models) are consistent with views that in older northeastern cities (such as New York and Boston) where white immigrants had historically controlled ethnic machine politics, blacks have less political power and stronger motivation to protest (Mink 1986).

Our results are extremely robust: addition of the other indicators did not alter the pattern of effects for dilapidated housing and black income. These results add to the growing doubt that poverty systematically causes unrest. They suggest that cities where blacks are doing relatively better are likely to experience more unrest than when blacks fall further behind whites in family income. Below we offer some explanations of why this might be true.

EFFECTS OF COMPETITION

Arguments from competition theory lead us to consider factors that intensify racial competition over jobs, housing, schools, and other opportunities for advancement. We hypothesized that an increase in the population at the same time that unemployment was high would intensify ethnic/racial conflict. This argument suggests that ethnic/racial competition intensifies in job markets where supply increases at the same time that demand for workers falls. We speculated that both generalized competition for jobs and competition within the minority niche might each affect the rate of race riots in cities.

Does competition add to our understanding of racial unrest? The positive and significant effects of both competition measures in Table 3 suggest that they do. Table 3 adds two indicators of competition. The first two columns measure the effect of rising competition overall in the labor force, while columns 3 and 4 specify the effect of competition within the minority community. We also compare the effects of these competition indicators in models that include those effects found to be important in the previous table. Thus the odd-numbered columns (1 and 3) of Table 3 include the percentage of dilapidated housing and the even-numbered columns (2 and 4) include the ratio of nonwhite to white income. In this way, the relative strength of competition processes can be judged alongside other significant factors.

The positive and significant coefficients for the two interaction effects of competition in all four columns strongly supports the competition argument. In the first two columns, overall competition in the labor force increases the rate of race riots, as expected. Furthermore, addition of this interaction effect does not diminish the significance of other covariates in the models. As in Table 2, the effects of percentage of dilapidated housing and the ratio of nonwhite to white income remain positive and significant where they are included in the model. Furthermore, both specifications of competition seem important. In the first two columns, overall competition raises the rate of race riots. In the second two columns we see that an increase in minority competition also increases the rate.

TABLE 3: Poisson Regression of Effects of Competition on Race Riots, 1953-1994

	(1)	(2)	(3)	(4)
Intercept	-6.29**	-8.38**	6.25	-8.40**
	(.904)	(.773)	(.902)	(.749)
Ln Population 1960	.636**	.657**	.648**	.668**
	(.053)	(.051)	(.050)	(.051)
Ln Percent nonwhite population 1960	.675**	.690**	.695**	.697**
	(.111)	(.116)	(.110)	(.109)
Percent change in population 1950-1960	-.853+	-.941*		
	(.429)	(.436)		
Percent change in nonwhite population 1950-60			-.710+	-.680
			(.440)	(.443)
Percent all males unemployed 1960	-3.20	-5.12		
	(4.971)	(5.16)		
Percent nonwhite males unemployed 1960			-4.50	-4.43
			(3.24)	(3.26)
Nonpartisan elections	.538**	.515*	.526*	.492*
	(.175)	(.175)	(.176)	(.1716)
Northeast (0,1)	.642**	.659**	.623**	.612**
	(.192)	(.190)	(.187)	(.187)
Percent dilapidated housing 1950	-5.11*		-5.03*	
	(1.59)		(1.63)	
Ratio of nonwhite/white median income 1960		2.55**		2.45**
		(.839)		(.799)
Increase in overall Competition 1950-60	14.6*	17.2*		
	(6.94)	(7.19)		
Increase in minority competition 1950-60			8.34*	8.43*
			(4.01)	(4.07)
Log likelihood	-34.56	-35.41	-35.31	-35.56

[a] Heteroskedastic-consistent standard errors are in parentheses.

+ $p < .10$ * $p < .05$ ** $p < .01$ (two-tailed tests)

In both cases, these effects of competition are net of the effects of population size, minority population, and nonwhite income.

How can we interpret the positive effect of competition? Using the same logic and calculations as described above, the three coefficients (for black unemployment, percentage change in nonwhite population, and the interaction between these two) from column 4 suggest that for cities with an average change in black population that experience a black unemployment rate that is one standard deviation higher than the mean rate (13.5% compared to 9.4%), the rate of race riots increases 30-fold. In other words, the interaction term together with main effects of unemployment and changing black population suggest that black unemployment intensifies the consequences of black in-migration and black in-migration intensifies the effects of black unemployment on unrest.

In Table 3 we see that population size and the proportion that is black in cities raises the likelihood of race riots, just as it did in Table 2. Just how big is the effect of minority nonwhite population? Recall that Spilerman found that black population size systematically diminished the effects of poverty and other characteristics. We reasoned that given the relationship found for changes in black population in the earlier period, it might be important to understanding how the proportion of blacks affects racial unrest independent of the in-migration. To isolate the effect of minority population level, we calculated the effect of black population size for cities experiencing an average influx of black migrants during the 1950-60 decade. Comparing cities at the lowest quartile of proportion that is black with the highest quartile, the coefficient of .675 would imply that the rate is increased 275% over this range. The effect of black population size remains an overwhelmingly potent one.

Once we control for population size, dilapidated housing, and competition, we fail to find any effects for most of the absolute and relative deprivation measures in our analysis over forty years.[22] In particular, low levels of black education, high levels of black participation in low-status occupations, as well as racial disparities in education, and disproportionate participation of blacks compared to whites in low-status occupations were unrelated to race riots.

EFFECTS OF POLITICAL STRUCTURE

Prior research suggested that certain changes in the political opportunity structure might be more (or less) conducive to racial tensions and rioting. Political process models echo this suggestion when they point out that several structural features of local government encouraged protest rates as well as other forms of mobilization (McAdam 1982). Spilerman (1971) found little support for these political indicators, despite the widespread assumptions that responsiveness of local government matters.

We investigate this debate as well. Following the political process line of argument, we include measures of nonpartisan elections, mayor-council form of government, proportion of population per council member, and percentage of council members elected at large. Although the zero-order correlations in column 2 of Table 1 suggest that these structural features of city politics were relevant to race riots, our regressions show little evidence in support of this claim. Only one of the hypothesized features of city politics had systematic

effects on race riots over this period (as well as for the entire period): that of nonpartisan elections.

The original argument for inclusion of nonpartisan elections was that cities with nonpartisan politics would be less responsive to minority grievances and demand. Furthermore, Spilerman argued that the unresponsiveness would be more relevant during the 1960s, during the peak of civil-rights activism (Spilerman 1971). Though Spilerman's analysis found that nonwhite population size dwarfed all aspects of city politics, we found that the presence of nonpartisan politics in a city contributes systematically to the occurrence of race riots (holding constant minority population size, city size, and other features of the community).

Our models attempt to distinguish the impact of region from the substantive effects of black population, minority poverty, and competition. This strategy proved important because of the regional concentration of black poverty in the South during this period and the high correlation of competition factors in the Northeast. By including all three regional dummy variables, some of these regional effects can be untangled from the economic competition measures. As in the sixties, the northeastern region produces more racial unrest, and this effect holds even when population composition, poverty, and city political structure are included in the model. Because southern location was significant in Spilerman's original analysis, we were surprised by the fact that southern location had little systematic effects in our analysis.

We would argue that the positive effect found for the northeastern region in all our models is highly consistent with our competition arguments. That is, it was primarily in the Northeast that the U.S. witnessed the beginning of the process of racial competition for the same jobs, housing, and other rewards, as black migrants moved northward (Wilson 1978). Up until recently, employment and wage rates for blacks were substantially better in the North and Northeast than they were in the South. This indicates that during the peak period of racial unrest in the U.S., racial deprivation was disproportionately southern and rural in character (Harris & Wilkins 1988).[23]

Discussion

We began this article with the puzzle that although deprivation theories have been empirically discredited, they remain popular explanations of past and present race riots. We were intrigued by the fact that deprivation explanations cannot account for the fact that race riots occur in places where blacks are relatively better off, or during periods when racial disparities are declining rather than rising. These contradictions shaped our research questions and motivated our reanalysis of the effects of racial poverty.

Farley and Allen (1989) and Wilson (1987) (among many others) have documented changes in minority class structure that provide some clues to understanding this puzzle. In their view, chances for upward mobility have improved for some African Americans but have declined for others at the very bottom of the ladder. The rise in class diversity within the black community is the result of several economic and demographic trends that diminished

opportunities for blacks in the urban working class. At the same time, for educated, middle-class African Americans, opportunities in jobs, suburban housing, and occupational mobility have apparently expanded (Massey & Denton 1993). An often-repeated consequence of these trends is the rapid growth of increasingly poor and black single-parent families that are isolated from city services, education, and employment training (Wilson 1987).

Those who argue that the racial tensions are caused by poverty have linked the growth of the underclass to the persistence of racial tensions and riots in American cities. In this view, the growing underclass constitutes a new and growing threat to harmonious race relations (Massey & Denton 1993). Some have reasoned that these factors could lead ethnic and racial tensions to rise exponentially (Harris & Wilkins 1988). Such an apocalyptic view suggests that American cities are particularly at risk for ethnic violence because of the rising ethnic and racial diversity from rapid rates of immigration and mobility.

Our results have implications for these scenarios. Our analysis does not justify the fears that violence will escalate as the underclass grows. This is because we found little evidence that black poverty or racial disparities in income, education, and occupation caused race riots. Furthermore, we do find that the effects of black poverty are generally opposite from what most deprivation theories would lead us to expect. Cities with more dilapidated housing and lower income for blacks (relative to whites) had significantly lower rates of racial unrest. Finally, we find that competition that is expressed as the contraction of job opportunities created by demographic pressures and unemployment significantly increases the rate of race riots.

These findings are consistent with the view that rates of racial protest rise when racial disadvantages are relatively low. In these cities, the opportunity for blacks to begin to compete alongside whites for similar jobs becomes possible. If the economy expands accordingly, conflict over jobs and housing can be averted. But under conditions of competition, the overall contraction of labor force demand combined with the rising supply of labor from migration intensifies interracial competition and conflict.

Assessing our results from a wide angle, we see that our analysis generally supports the thrust of earlier work from earlier periods, but with a competition twist added to it. Thus we diverge from earlier studies in arguing that it is not the case that deprivation is irrelevant to race riots. Rather, we found there are some indications that cities where blacks are less economically deprived have higher rates of race riots. We saw hints of this finding in earlier work by Spilerman and Lieberson and we found it intriguing. Following the example of the best of the prior research that avoided sample selection bias, we analyzed city characteristics from locations that had no rioting as well as those with multiple riots. Corresponding to Spilerman's and Lieberson's results, we also find that few indicators of poverty, deprivation, or black/white gaps in achievement affect the likelihood of race riots. But we also find one new result: several indicators of rising competition do affect the rate of race riots and they do so significantly.

Despite the popularity of deprivation theories, the empirical analysis shown here suggests that race riots arise where minority deprivation is diminishing and where competition is rising. When migration and immigration create tight

labor markets, black unemployment contributes significantly to unrest in cities. At the same time, our results show that an influx of minorities places additional pressure on the minority community. Our results suggest that one strategy for understanding race riots can be found by exploring the forces that elevate levels of racial competition.

Summary and Conclusions

Our results suggest some avenues of research on the future of race relations. If poverty were directly related to unrest, then the stability of race relations in America's cities would indeed be in peril. Yet our results suggest another more complicated scenario. The pattern of results suggests that poverty alone does not spark racial violence. Instead, competition for increasingly scarce resources such as jobs and housing during periods of increasing population growth produces racial unrest. The relationships uncovered in this research suggest that racial violence is less likely when minority in-migration and immigration are matched by an expanding labor market. For those interested in reducing ethnic/racial conflict, the policy implications are straightforward.

The analysis draws attention to the role of changes in the opportunity structure and how those changes can have both positive and negative effects on race relations. The arguments offered here suggest that both rising and declining economic opportunity can raise rates of racial violence if resources are limited and employment (and other) opportunities remain static. Competition intensifies only to the extent that demands are placed on a closed system of limited resources.

Notes

1. These data on riots are part of a larger data set (1954-1993) on all forms of ethnic/racial collective action that were collected at the SMSA (Standard Metropolitan Statistical Area) level. The analysis first required disaggregation of the SMSA-level data to match city administrative units from Spilerman's list of cities (personal communication). The riot count for 204 cities that were analyzed can be found in Appendix C. Further information on the location of SMSAs and central city units can be obtained from the first author.

2. For examples, see research on white antibusing violence in Useem (1980); ethnic violence in Northern Ireland, in White (1993), and ethnic conflict in the Soviet Union, in Tilly (1991).

3. For national time-series analyses of race riots that extend through the late 1970s, see Kelly and Isaac (1984).

4. The competition arguments developed here also rely on human ecology theories, initially developed by Park (1949). For applications of these theories to instances of ethnic and collective action, see Blalock (1967), Jenkins and Shock (1991), Lieberson (1961), Medrano (1994), Ragin (1977, 1979), See and Wilson (1988), and Wilson (1978). See also the wide range of empirical studies using this perspective in Belanger and Pinard (1991), Olzak and Nagel (1986), Olzak and West (1991), Tilly (1991).

5. Human ecologists emphasize that conflict and competition do not always coincide. Niche separation or ethnic dominance over subordinate groups can also result from competition (Barth 1969).

6. For historical evidence supporting this point, see Senechal (1990), and Tuttle (1970).

7. We began our analysis with race riots from 318 SMSAs during this period and subsequently coded whether a given race riot occurred in a center city of an SMSA or in the suburban or surrounding area. Race riots that were in one of two center cities (for example, in San Francisco rather than in Oakland in the San Francisco-Oakland SMSA) were identified. We then recoded our events to indicate a center city code to coincide with Spilerman's center city code. For comparability between our event data and 1960 census data on cities, we created an event history data set for (1) center cities of SMSAs that existed in (2) 1960 and (3) remained as part of the SMSAs in 1980 or (4) had a population of over 100,000 in 1960. This means that we have virtually all center cities for the SMSAs that existed as SMSAs in 1960. We have several sets of twin center cities that were merged into one SMSA in 1980 (such as Dallas-Fort Worth, San Francisco-Oakland). The analysis treats each of these twin center cities as a separate unit of analysis. The total number of cities for which we have data on covariates and events is 204 center cities (see Appendix C).

8. We exclude here all large-scale unruly events that involved predominately white residents, as in the case of antibusing riots.

9. One reviewer preferred the term *commodity riots* to *race riots* for the present paper, citing Janowitz (1968) and McPhail (1994). In our view, this category is imprecise, because, although violence against property and looting often occurs as part of the actions in a riot, including race riots, it is not the defining feature of race riots. In our view, assuming that all race riots since 1963 are "commodity" riots would lead to confusion of contemporary race riots with other forms of property riots: soccer riots, grain and food riots, looting from brownouts, Luddite violence, and other property-related violence. In making this decision, we emphasize that we are categorizing the event, not the participants. Thus, our coding rule does not require that all participants be from one race. Our coding scheme is available from the first author upon request, as is the list of riot dates.

10. Here we draw a distinction between event analysis and content analysis. Gathering event data for use in event history analysis (used in estimating causal models elaborated here) necessitates reliable information on the dates, duration, and nature of the conflict, rather than information on the specific word counts or analysis of the reporter's characterizations of participants and their actions. Such questions are more appropriate for applying content analysis to specific news reports, papers, or editors.

11. Recent reviews and analyses of newspaper sources of data on events can be found in Franzosi (1987), McCarthy et al. (1996), and Olzak (1992). Empirical analyses of the sources of potential bias problems have been conducted by McCarthy et al. (1996) and Snyder and Kelly (1977).

12. We thank Dan Myers for challenging our use of mainly nonvarying covariates from 1960. He points out that there have been significant changes in the racial composition (or diversity) of these cities across the thirty-five years of this study. Many believe these demographic shifts have exacerbated racial tensions, at least in the recent Los Angeles riots (Baldassare 1994). Ideally, we would have liked to include changing levels of these covariates, but we encountered several obstacles. In particular, the definitions of "black" shifts from nonwhite in 1950 (and 1960) to black in 1970. Even more troubling, the category of non-Hispanic black is distinguished by the census in 1980 and 1990 (and nonwhite ceases to be applied) (Massey & Denton 1988, 1989). Our observations on racial unrest begin in 1960, and so we are constrained to use the antiquated categories of nonwhite for most of our race-specific covariates for this analysis of center cities (but see Olzak, Shanahan & McEneaney 1996).

13. Spilerman's results led us to believe that an absence of effects found for black poverty would be as theoretically interesting and relevant as would any positive or negative effects of poverty. Thus our strategy was to include as many of the indicators as possible. However, because of the multicollinearity between absolute and relative deprivation measures (which are linear combinations of each other), it was not possible to produce analyses of models that include all the indicators described in this section. To avoid estimation problems caused by inflated standard errors, our tables report results for those indicators that survived the log-likelihood test for significant contribution to the overall fit of the model. Concretely, this means that we added each deprivation variable separately in models that included population and

APPENDIX A: Descriptive Statistics for Community Characteristics

Variable	N	Mean	Std. Dev.	Min.	Max.
Ln Population 1960	204	11.82	.97	10.24	15.87
Ln Percent nonwhite 1960	204	-2.17	.92	-4.51	-.60
Change percent total population 1950-60	204	.32	.58	-.17	3.68
Change percent nonwhite population 1950-60	204	.69	.70	-.88	3.98
Percent total male unemployment 1960	204	.05	.02	.01	.29
Percent nonwhite male unemployment 1960	204	.09	.04	.005	.29
Nonpartisan elections 1960	204	.70	.46	.00	1.00
Northeast	204	.16	.36	.00	1.00
Percent dilapidated housing units 1950	204	.08	.06	.003	.317
Relative income 1960 (nonwhite median family income/ white figure)	204	.61	.11	.33	.92
Relative education 1960 (nonwhite median education/ white figure)	204	.78	.13	.34	1.11
Relative occupation 1960 (percent nonwhite in low-status jobs/ white figure)	204	2.04	.71	.29	4.42
Overall competition 1960 (ln population * total unemployment)	204	.01	.03	-.01	.23
Minority competition 1960 (ln percent nonwhite * nonwhite unemployment)	204	.07	.08	-.13	.45

political measures. Below we discuss the results for all these covariates in the text so that readers can compare the results with Spilerman's analyses.

14. For another sociological application of Poisson regression applied to another form of racial violence, see Beck and Tolnay (1994).

15. Appendixes A and B present the means, standard deviations, and correlations of the key covariates in Tables 2 and 3. While multicollinearity is often present in analyses of poverty and racial composition measures, it is interesting that the correlations among the natural log of percentage nonwhite with the other measures of "underclass poverty" (nonwhite unemployment rate, percentage of dilapidated housing, and relative income and education) are so low for these 204 cities. The deprivation and relative deprivation indicators are of course highly collinear with each other. In most cases, this required that we enter them singly in each model.

16. We agree that it would have been preferable to distinguish race/ethnic categories more precisely. Current research using data from the 1980s and 1990s now distinguishes Hispanics from non-Hispanics, by race. For examples, see Massey and Denton (1989, 1993). Unfortunately, these key distinctions are unavailable for earlier decades. Concretely, this means that Spilerman's original data is limited by the prevailing racial category of nonwhite, which we now know contains some (unknown) percentage of Hispanics, Asians, and Native Americans, who classified themselves as nonwhite.

17. We examined all the covariates of deprivation and relative deprivation (from Table 1) and found that only the effects of dilapidated housing and the ratio of nonwhite to white family income were significant. For this reason we focus on only those measures that had systematic effects.

18. The presence of multicollinearity required that we enter the covariates independently in most cases. This was particularly true for the socioeconomic indicators of nonwhite education and income (the Pearson coefficient is .70). Although we investigated the effects of all the covariates listed in Table 1, only those reported below had any systematic effects once population size and minority composition were taken into effect. In models analyzed (but not shown), we found no significant effects for the poverty indicators: median black education, black participation in low-status occupations, and black unemployment level (by itself) apparently has little systematic impact on racial unrest.

19. Table 2 reports a likelihood ratio to compare the goodness of fit for nested models. If the null hypothesis holds, then the difference between the likelihood ratios has a chi-square distribution with degrees of freedom equal to the number of constraints that have been freed. For example, to examine whether the addition of dilapidated housing improves the fit, compare the log-likelihood of -37.54 in column 2 to -42.04 in column 2. Twice the difference is 9.0, which is significant at the .01 level (with one degree of freedom).

20. Since this effect also plays a central role in Spilerman's (1970a) conclusions, it may be useful to explain how this effect is calculated over the observed range of black population size. For example, the coefficient for the log of population size in 1960 is .712 in column 1. This parameter suggests that when city size is one standard deviation higher than the mean, race riots nearly double. This effect is calculated by taking the exponent of the product of the effect of black population size, .712, at one standard deviation above the mean and comparing this effect to that found at the mean for this variable.

21. For further explanation of interpretation of Poisson regression coefficients, see Liao (1994:72-74).

22. We remind readers that we do not have direct measures of perceived levels of deprivation or aspirations, but rather, like Spilerman, we use only indirect indicators: black/white median income, education, and low-status occupations.

23. One of the anonymous reviewers asked if we did analyses to discover if competition levels differed in Chicago and New York (the two cities with the highest rate of race riots in our data). If we understand this comment, there is an alternative hypothesis underlying this comment worth investigating. Following this logic, we ran the same regressions as we report in Tables 2 and 3, omitting these two "high riot-count" cities. Our competition results remained the same, which was reassuring. In general, however, our modeling strategy is to try to capture the effects of "outlier" cities by including the appropriate covariates rather than by including specific city dummy variables (e.g., including a dummy variable for both Chicago and New

APPENDIX B: Correlations for Community Characteristics[a]

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Ln Population 1960													
2. Ln Percent nonwhite 1960	.15												
	.03												
3. Change percent total population 1950-60	-.10	-.12											
	.14	.08											
4. Change percent nonwhite population 1950-60	.03	-.35	.53										
	.62	.00	.00										
5. Percent total male unemployment 1960	.06	-.11	-.11	-.04									
	.35	.90	.10	.52									
6. Percent nonwhite male .06 unemployment 1960	-.12	-.18	.05	.41									
	.38	.07	.01	.47	.00								
7. Nonpartisan elections 1960	-.06	-.04	.18	.08	-.02	-.13							
	.41	.57	.01	.23	.72	.07							
8. Northeast	.14	-.10	-.19	.06	.15	.18	-.39						
	.04	.16	.00	.35	.03	.01	.00						
9. Percent dilapidated housing units 1950	-.24	.46	.14	-.28	-.14	-.28	.04	-.30					
	.00	.00	.04	.00	.05	.00	.54	.00					
10. Relative income 1960 (nonwhite median family income/white figure)	.22	.46	-.13	.30	.13	.15	-.01	.28	-.63				
	.00	.00	.07	.00	.06	.03	.90	.00	.00				
11. Relative education 1960 (nonwhite median education/ white figure)	.34	-.48	-.19	.23	.20	.15	-.02	.41	-.56	.71			
	.00	.00	.00	.00	.00	.04	.37	.00	.00	.00			
12. Relative occupation 1960 (percent nonwhite in low-status jobs/white figure)	-.25	-.65	.18	.19	-.27	.12	.09	-.13	-.22	.03	.15		
	.00	.00	.00	.00	.70	.08	.19	.05	.00	.71	.03		
13. Overall competition 1960 (Ln population * total unemployment)	-.09	-.14	.86	.49	.29	-.13	.16	-.21	.11	-.08	-.14	.17	
	.20	.05	.00	.00	.00	.06	.02	.00	.11	.24	.05	.02	
14. Minority competition 1960 (Ln nonwhite * nonwhite unemployment)	.04	-.32	.35	.86	.10	.44	.02	.12	-.33	.27	.22	.22	.35
	.60	.00	.00	.00	.16	.00	.79	.08	.00	.00	.00	.00	.00

[a] Probability appears below estimate.

APPENDIX C: Number of Black Instigated Riots in Each Community[a]

Abilene	0	Fayetteville	0	Midland	0	Santa Barbara	0
Akron	0	Flint	1	Milwaukee	4	Savannah	0
Albany, Ga.	1	Fort Smith	0	Minneapolis	2	Seattle	2
Albany, N.Y.	1	Fort Wayne	0	Mobile	0	Shreveport	0
Albuquerque	0	Ft. Lauderdale	4	Monroe	0	Sioux City	0
Amarillo	0	Ft. Worth	0	Montgomery	2	South Bend	1
Anderson	0	Fresno	1	Muncie	0	Spokane	0
Ann Arbor	0	Gainsville	0	Muskegon	1	Springfield, Ill.	0
Anniston	0	Galveston	0	Nashville	2	Springfield, Ohio	0
Asheville	0	Gary	1	New Bedford	2	Springfield, Mass.	0
Atlanta	5	Grand Rapids	2	New Haven	1	Springfield, Mo.	0
Atlantic	2	Great Falls	0	New Orleans	0	St. Joseph	0
Augusta	2	Greensboro	2	New York	23	St. Louis	4
Austin	0	Greenville	0	Newark	4	St. Petersburg	0
Baltimore	3	Hamilton	2	Newport News	0	Steubenville	0
Baton Rouge	4	Harrisburg	1	Norfolk	0	Stockton	0
Beaumont	0	Hartford	5	Oakland	0	Syracuse	1
Bilouxi	0	Houston	1	Odessa	0	Tacoma	0
Binghamton	0	Huntington	0	Ogden	0	Tallahassee	0
Birmingham	6	Huntsville	0	Oklahoma	0	Tampa	5
Boston	1	Jackson, Miss.	0	Omaha	5	Terrehaute	0
Bridgeport	1	Jackson, Mich.	2	Orlando	0	Texarkana	1
Buffalo	2	Jacksonville	0	Owensboro	1	Toledo	2
Canton	0	Jersey City	6	Oxnard	0	Topeka	0
Cedar Rapids	0	Johnson City	2	Patterson	5	Trenton	1
Champaign	0	Johnstown	0	Pensacola	0	Tucson	0
Charleston	0	Kalamazoo	0	Peoria	0	Tulsa	0
Charleston	0	Kansas City	3	Petersburg	0	Tuscaloosa	0
Charlotte	1	Kenosha	1	Philadelphia	5	Tyler	0
Chattanooga	0	Knoxville	0	Phoenix	1	Utica	0
Chicago	19	Lafayette	0	Pine Bluff	0	Vallejo	2
Cincinnati	4	Lake Charlotte	0	Pittsburgh	1	Vineland	0
Cleveland	4	Lancaster	0	Portland	1	Waco	0
Colorado	0	Lansing	0	Providence	2	Washington, D.C.	4
Columbia	0	Las Vegas	2	Pueblo	0	Waterbury	2
Columbia	0	Lawton	0	Racine	0	Waterloo	2
Columbus	2	Lexington	0	Raleigh	0	West Palm	0
Columbus	1	Lima	1	Reading	0	Wheeling	0
Corpus Christi	0	Lincoln, Nebr.	0	Reno	0	Wichita	2
Dallas	0	Little Rock	1	Richmond	1	Wichita Falls	0
Davenport	0	Long Beach	0	Riverside	2	Wilmington	2
Dayton	3	Loraine	0	Roanoke	0	Wilmington	3
Decatur	1	Los Angeles	5	Rochester	4	Winston	0
Denver	1	Louisville	1	Rockford	0	Worcester	0
Des Moines	1	Lubbock	0	Saginaw	1	Yakima	0
Detroit	5	Lynchburg	0	Salinas	0	York	3
Duluth	0	Macon	0	Salt Lake	0	Youngstown	4
Durham	0	Madison	0	San Angelo	0		
El Paso	0	Mansfield	0	San Diego	3		
Erie	1	Memphis	3	San Francisco	6		
Evansville	1	Miami	9	San Jose	0		

(N = 204)

References

Aalen, Odd O. 1978. "Nonparametric Inference for a Family of Counting Processes." *Annals of Statistics* 6:701-26.

Aiken, Michael, and Robert Alford. 1970. "Community Structure and Innovation: The Case of Urban Renewal." *American Sociological Review* 35:650-65.

Baldassare, Mark. 1994. *The Los Angeles Riots: Lessons for the Urban Future.* Westview Press.

Barron, David N. 1992. "The Analysis of Count Data: Overdispersion and Autocorrelation." Pp. 179-220 in *Sociological Methodology 1992,* edited by Peter V. Marsden. Blackwell.

Barth, Fredrik. 1969. *Ethnic Groups and Boundaries.* Little, Brown.

Beck, E.M., and Stewart E. Tolnay. 1994. "Violence toward African-Americans in the Era of the White Lynch Mob." Pp. 121-44 in *Ethnicity, Race, and Crime: Perspectives across Time and Place,* edited by Darnell F. Hawkins. New York State University Press.

Blalock, Hubert M., Jr. 1957. "Percent Non-White and Discrimination in the South." *American Sociological Review* 22:677-82.

____. 1967. *Toward a Theory of Minority-Group Relations.* Wiley.

Blauner, Robert. 1971. *Racial Oppression in America.* Harper & Row.

Bloombaum, Milton. 1968. "The Conditions Underlying Race Riots As Portrayed by Multidimensional Scalogram Analysis: A Reanalysis of Lieberson and Silverman's Data." *American Sociological Review* 33:76-95.

Bonacich, Edna. 1972. "A Theory of Ethnic Antagonism: The Split Labor Market." *American Sociological Review* 37:547-59.

Boskin, Joseph. 1976. *Urban Racial Violence in the Twentieth Century.* Glencoe Press.

Cameron, A.C., and P.K. Trivedi. 1986. "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests." *Journal of Applied Econometrics* 1:29-53.

Chicago Commission on Race Relations, The. 1922. *The Negro in Chicago: A Study of Race Relations and a Race Riot.* University of Chicago Press.

Coser, Lewis A. (ed.). 1972. "Collective Violence and Civil Conflict." Special Issue of *Journal of Social Issues* 28.

Downes, Bryan T. 1969. "Social and Political Characteristics of Riot Cities: A Comparative Study." Pp. 427-43 in *Blacks in the United States,* edited by Norval Glenn and Charles M. Bonjean. Chandler.

Economist, The. 1992. "America's Cities: Doomed to Burn?" 9 May, 1992, pp. 21-24.

Eisinger, Peter K. 1973. "The Conditions of Protest Behavior in American Cities." *American Political Science Review* 67:11-28.

Farley, Reynolds, and Walter R. Allen. 1989. *The Color Line and the Quality of Life in America.* Oxford Press.

Feagin, Joe R., and Harlan Hahn. 1973. *Ghetto Revolts.* Macmillan.

Franzosi, Roberto. 1987. "The Press as a Source of Socio-Historical Data: Issues in the Methodology of Data Collection from Newspapers." *Historical Methods* 20:5-16.

Graham, Hugh D., and Ted R. Gurr. 1969. *Violence in America.* Signet.

Grimshaw, Allen D. 1960. "Urban Racial Violence in the United States: Changing Ecological Considerations." *American Journal of Sociology* 66:109-19.

____. 1969. *Racial Violence in the United States.* Aldine.

Gurr, Ted Robert. 1972. "The Calculus of Civil Conflict." *The Journal of Social Issues* 28:27-48.

Harris, Fred R., and Roger W. Wilkins. 1988. *Quiet Riots: Race and Poverty in the United States.* Pantheon.

Higgs, Robert. 1977. *Competition and Coercion: Blacks in the American Economy, 1865-1914.* Cambridge University Press.

Janowitz, Morris. 1968. *Social Control of Escalated Riots.* University of Chicago Press.

Jencks, Christopher, and Paul E. Peterson (eds.). 1991. *The Urban Underclass.* The Brookings Institute.

Jenkins, J. Craig. 1983. "Resource Mobilization Theory and the Study of Social Movements." Pp. 527-53 in *Annual Review of Sociology*. Vol. 9. Annual Reviews.

Kelly, William R., and Larry Isaac. 1984. "The Rise and Fall of Urban Racial Violence in the U.S.: 1948-1979." *Research in Social Movements, Conflict, and Change* 7:203-33.

Ladner, R.A., B.J. Schwartz, S.J. Roker, and L.S. Titterand. 1981. "The Miami Riots of 1980: Antecedent Conditions, Community Responses and Participant Characteristics." *Research in Social Movements, Conflicts, and Change* 4:171-214.

Lee, Alfred McClung, and Norman D. Humphrey. 1943. *Race Riots.* Dryden Press.

Liao, Tim. 1994. *Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models.* Sage.

Lieberson, Stanley. 1961. "A Societal Theory of Race Relations." *American Sociological Review* 26:902-908.

Lieberson, Stanley, and Arnold R. Silverman. 1965. "The Precipitants and Underlying Conditions of Race Riots." *American Sociological Review* 30:887-98.

Massey, Douglas S., and Nancy A. Denton. 1988. "The Dimensions of Residential Segregation." *Social Forces* 67:281-315.

____. 1989. "Hypersegregation in U.S. Metropolitan Areas: Black and Hispanic Segregation along Five Dimensions." *Demography* 26:373-92.

____. 1993. *American Apartheid: Segregation and the Making of the Underclass.* Harvard University Press.

McAdam, Doug. 1982. *Political Process and the Development of Black Insurgency.* University of Chicago Press.

McAdam, Doug, John D. McCarthy, and Mayer Zald. 1988. "Social Movements." Pp. 695-738 in *Handbook of Sociology*, edited by Neil J. Smelser. Sage.

McCarthy, John D., Clark McPhail, and Jackie Smith. 1996. "Images of Protest: Estimating Selection Bias in Media Coverage of Washington Demonstrations 1982, 1991." *American Sociological Review.* In press.

McPhail, Clark. 1994. "The Dark Side of Purpose: Individual and Collective Violence in Riots." *Sociological Quarterly* 35:1-32.

Medrano, Juan Díez. 1994. "The Effects of Ethnic Segregation and Ethnic Competition on Political Mobilization in the Basque Country, 1988." *American Sociological Review* 59:873-89.

Mink, Gwendolyn. 1986. *Old Labor and New Immigrants in American Political Development.* Cornell University Press.

Morgan, William R., and Terry N. Clark. 1973. "The Causes of Racial Disorders: A Grievance-Level Explanation." *American Sociological Review* 38:611-24.

Morris, Aldon. 1984. *The Origins of the Civil Rights Movement: Black Communities Organizing for Change.* Free Press.

Myrdal, Gunnar. 1944. *An American Dilemma: The Negro Problem and Modern Democracy.* Harper & Row.

Oliver, Melvin L., and James H. Johnson Jr. 1984. "Inter-Ethnic Conflict in an Urban Ghetto." *Research in Social Movements, Conflict, and Change* 6:57-94.

Olzak, Susan. 1989. "Analysis of Events in Studies of Collective Action." Pp. 119-41 in *Annual Review of Sociology*. Vol. 15, edited by W. Richard Scott and Judith Blake. Annual Reviews.

____. 1992. *The Dynamics of Ethnic Competition and Conflict.* Stanford University Press.

Olzak, Susan, and Lisa Brooke. 1995. "Investigation of Bias in Newspaper Reporting of Race Riots in the U.S." Unpublished manuscript, Stanford University.

Olzak, Susan, and Elizabeth West. 1989. "Ethnic Conflicts and the New Immigrants, 1965-1985." Paper presented at the annual meeting of the American Sociological Association, San Francisco.

____. 1990. "Ethnic Conflict and Protest in the United States, 1954-Present: Research Design and Codebook." Unpublished Codebook Manual. Stanford University.

____. 1991. "Ethnic Conflict and the Rise and Fall of Ethnic Newspapers." *American Sociological Review* 56:458-76.

Olzak, Susan, Suzanne Shanahan, and Elizabeth H. McEneaney. 1996. "Poverty, Segregation, and Race Riots, 1960-1993." *American Sociological Review*. In press.

Paige, Jeffrey M. 1971. "Political Orientation and Riot Participation." *American Sociological Review* 36:810-20.

Park, Robert E. 1949. *Race and Culture*. Free Press.

Petersen, Trond. 1991. "Time-Aggregation Bias in Continuous-Time-Hazard-Rate Models." *Sociological Methodology* 21:263-90.

Piven, Frances Fox, and Richard A. Cloward. 1977. *Poor People's Movements*. Vintage.

Rossi, Peter H. 1973. *Ghetto Revolts*. Transaction.

Senechal, Roberta. 1990. *The Sociogenesis of a Race Riot*. University of Illinois Press.

Snyder, David, and William R. Kelly. 1977. "Conflict Intensity, Media Sensitivity and the Validity of Newspaper Data." *American Sociological Review* 42:105-23.

Spilerman, Seymour. 1970a. "The Causes of Racial Disturbances: A Comparison of Alternative Explanations." *American Sociological Review* 35:627-49.

____. 1970b. "Comment on Wanderer's Article on Riot Severity and Its Correlates." *American Journal of Sociology* 75:556-59.

____. 1971. "The Causes of Racial Disturbances: Test of an Explanation." *American Sociological Review* 35:427-42.

____. 1976. "Structural Characteristics of Cities and the Severity of Racial Disorders." *American Sociological Review* 41:771-93.

Tilly, Charles. 1978. *From Mobilization to Revolution*. Random House.

____. 1991. "Ethnic Conflict in the Soviet Union." *Theory and Society* 20:569-81.

Tolnay, Stewart E., and E.M. Beck. 1995. *A Festival of Violence*. University of Illinois Press.

Tuttle, William M. 1970. *Race Riot: Chicago in the Red Summer of 1919*. Athenaeum.

U.S. Bureau of the Census. 1950. *Characteristics of the Population*. Vol. 1. Government Printing Office.

U.S. National Advisory Commission on Civil Disorders. 1968. *Report of the National Advisory Commission on Civil Disorders*. Bantam Books.

Useem, Bert. 1980. "Solidarity Model, Breakdown Model and the Boston Anti-Busing Movement." *American Sociological Review* 45:357-60.

Wanderer, James J. 1969. "An Index of Riot Severity and Some Correlates." *American Journal of Sociology* 74:500-505.

White, Robert W. 1993. "On Measuring Political Violence: Northern Ireland, 1969 to 1980." *American Sociological Review* 58:575-85.

Wilson, William J. 1981. *The Declining Significance of Race*. 2d ed. University of Chicago Press.

____. 1987. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. University of Chicago Press.

Labor Force Transitions among Older African American and White Men*

JEFFREY A. BURR, *State University of New York at Buffalo*
MICHAEL P. MASSAGLI, *University of Massachusetts-Boston*
JAN E. MUTCHLER, *State University of New York at Buffalo*
AMY M. PIENTA, *State University of New York at Buffalo*

Abstract

Using a life-course/opportunity-cost framework, we study racial differences in labor force behavior among African American and white men aged 55 to 69. A multifaceted measure of labor force behavior is examined within a longitudinal framework. We perform the analyses with a merged sample of the 1984 and 1985 Survey of Income and Program Participation, and we find that the most stable status is not working, followed by full-time, part-time, and unemployed statuses. Results from multivariate logistic regression change models show race-specific effects of age, health, and not-working status on several labor force status and attrition contrasts. Researchers have much to gain by continuing to consider racial differences in late-life labor force behavior and by focusing on contemporaneous and lagged measures of life-course variables.

Given the increasing diversity of the American labor force, the sociological value of traditional definitions of retirement is decreasing because labor force withdrawal has become much more complex than moving directly from full-time attachment to work to full cessation of work (see Atchley 1982). For a substantial number of older persons, labor force exit is a process and not a single, irreversible event. In addition, the process varies with sociodemographic characteristics such as race, ethnicity, and gender. In fact, some observers argue that retirement is more a white phenomenon than an African American one (Hayward, Friedman & Chen 1996).

For these reasons and others, such as cohort progression and change in state welfare and pension policies (see Henretta 1992), the popular image of a white,

** An earlier version of this article was presented at the 1994 annual meetings of the American Statistical Association (Toronto). We thank Ann M. Kremers for editorial assistance with this paper and the anonymous reviewers who provided insightful comments. The research reported in this article was supported by a grant from the AARP Andrus Foundation. Direct correspondence to Jeffrey A. Burr, 430 Park Hall, Department of Sociology, State University of New York at Buffalo, Buffalo, NY 14260.*

 Social Forces, March 1996, 74(3):963-982

middle-class male voluntarily moving into retirement at age 65 (or earlier) after years of service to a company is fast becoming an anachronism for a substantial portion of American workers. Thus, the question is whether a classic model of retirement behavior (e.g., Kotlikoff & Wise 1989; Quinn, Burkhauser & Myers 1990) is too narrowly defined for the heterogenous U.S. labor pool.

We argue that alternative approaches to the classic model of retirement are needed to help better understand race differences in labor force transitions. There are many reasons to expect differences between African Americans and whites. Compared to whites, African Americans on average have worse health, less education, fewer economic resources, and suffer the negative effects of discrimination, generally having access to the least desirable jobs. The impact of these reduced life chances accumulate over the life course so that by the time African Americans approach the traditional retirement age their opportunities for an "orderly" exit are reduced. As a result, African Americans have a weaker attachment to the labor force later in life and lower opportunity costs than whites. These issues lead us to believe that African American and white men will make different decisions about work in later life.

This article examines these issues by comparing late-life labor force transitions between African American and white men in the U.S. during the mid-1980s, using pooled data from the 1984 and 1985 panels of the Survey of Income and Program Participation. In order to accomplish this, we establish a measure of late-life labor force behavior that allows us to evaluate reductions in number of work hours, complete withdrawal from the labor market, reentry into the labor force, and other transitions.

Race and Transitions in Late-Life Labor Force Status

A more refined model of "retirement" must first take into account the myriad forms and causes of retirement. One of these, early retirement, is far more common today than in previous decades. For some older persons, earlier retirement combined with increasing longevity provides a variety of opportunities for reentry into the labor force (Anderson, Burkhauser & Butler 1984; Beck 1985; Hayward, Hardy & Liu 1994). Reentry may mean part-time or full-time employment, often in occupations similar to the individual's primary occupation before retirement (Hayward, Hardy & Liu 1994).

Many older persons do not voluntarily retire but rather fade from labor force activity as a result of being laid off or fired. Some portion of involuntary exits from the labor force in later life are brought on by protracted spells of illness or disability (e.g., Chirikos & Nestel 1989; Hayward & Grady 1990). Under such circumstances, the maintenance of a full-time occupation becomes more difficult, if not impossible. Yet these circumstances may be reversible. Full or partial recovery from an illness or disability may allow an older person to reenter the job market, at least at a reduced level of participation, and likely in a different occupation. Still, whatever the cause of involuntary retirement, older unemployed workers face two barriers to new employment: ageism and the possession of obsolete skills (e.g., DeViney & O'Rand 1988; Lazcko 1989).

A retirement model must also account for labor force behavior among diverse sociodemographic groups. A model based on the behavior of white males may not be entirely satisfactory for explaining similar activities among women, members of other racial and ethnic groups, and the chronically poor or ill (e.g., Gibson 1987, 1991a; Jackson & Gibson 1985; Pienta, Burr & Mutchler 1994; Szinovacz 1985). Weak attachments to the labor market early in life, poor or unstable health, poor earnings opportunities, and the absence of pension income are potential sources of work instability in later life, especially among African Americans (Parnes & Nestel 1981). Those who experience less than optimal opportunity and reward structures over the life course possess labor market histories, economic profiles, and health/disability statuses that preclude a classic transition into permanent retirement.

Although research continues to document patterns of racial diversity in labor force participation in the U.S. (see Smith 1986; U.S. Bureau of Labor Statistics 1994; Welch 1990), much of our knowledge of late-life labor force patterns is still based on white males who retired in the 1970s and early 1980s. As recently as 1985, Jackson and Gibson (1985) state that "in spite of the need, virtually no research has focused on the work and retirement of older blacks" (194). A small number of social scientists have heeded this call. Recent studies comparing racial groups near retirement age agree that some important differences by race are apparent (Gibson 1987, 1991a; Gohmann 1990; Gustman & Steinmeier 1986; Hayward, Friedman & Chen 1996). These differences suggest to us that further study of the work patterns and transition behavior of older African Americans can help in the design of social policy as well as in the development of theory about labor force behavior within heterogeneous populations.

For example, Jackson and Gibson (1985) argue that the term *retirement* may have little meaning for many African Americans, who do not experience a marked decrease in work time in later life, but rather display continuous instability in work behavior over the life course until death or full disability. Gibson (1987) argues further that many older African Americans occupy an increasingly common labor force category that she terms the *unretired retired*. In her view, this phenomenon fits hand-in-hand with upward trends in physical disability and declines in labor force participation.

Scholars from different disciplines have discussed some of the reasons for the apparent correlation between increasing physical disability and decreasing labor force participation among older African Americans. Gibson (1987, 1991a) suggests that African Americans' relatively frequent adoption of the sick role stems from the increasing availability of disability benefits and the relatively low economic value of available labor market compensation and benefit levels. This in turn lowers their likelihood of claiming subjective retirement and increases their likelihood of subjective disability. Social and psychological benefits have also been attributed to this phenomenon.

Similar explanations have been suggested by others (Chirikos & Nestel 1985; Hayward, Friedman & Chen 1996; Parsons 1980). For example, Parsons (1980) finds a significant race effect when evaluating the impact of replacement income from Social Security disability benefits on labor force participation.[1] African Americans are more likely than their white counterparts to elect not to

participate in the labor force in the presence of disability benefits. However, the race effect disappears when Parsons controls for state-level differences in welfare benefits.

Hayward, Friedman, and Chen (1996), employing panel data from the National Longitudinal Survey of Older Men, find that the poorer health of African Americans compared with that of whites contributes to earlier exits from the labor force. In general, the empirical evidence about the existence and nature of race differences in the relationship between health and labor force participation patterns is mixed: Some researchers argue that race differences are based on subjective evaluations and others argue that the race differences are based on objective differences in health status. We expect that both the subjective evaluation of poor health/disability status and the actual race differences in health observed by others play important roles in the decision to work.[2]

Our interest in describing racial differences in labor force transitions is not limited to health/disability issues. In this study, we also incorporate economic characteristics of the individual. Since the economic histories of older African Americans are known to be poorer than those of older whites, we expect income in later life to play an important part in deciding how much labor, if any, to engage in. For example, older persons with fewer economic resources are often motivated to work part-time after withdrawing from a full-time position. In addition, persons who are not working at one point in time may need to reenter the labor market to make up for short-falls in pension income, if it exists at all.

Socioeconomic class may also account for differences in labor force instability between racial groups. Persons who are satisfied with their occupations are motivated to maintain longer attachments to the labor force, suggesting greater stability. Family status also impacts work decisions, and race differences in family stability and structure are well documented. Generally, the literature shows that married persons frequently retire jointly and may leave the work force earlier (Henretta & O'Rand 1983). However, little work has been reported to guide our expectations about racial differences in marital status effects on late-life labor force participation.

Age is also central to the decision process. Even within relatively narrow age ranges, age may play a role in racial differences in labor force behavior. For the youngest of the older persons considered in most retirement research, there may be institutional barriers to early withdrawal (e.g., eligibility for Social Security). On the other hand, institutional rules for private pension receipt may encourage early withdrawal for persons approaching traditional retirement ages. Since African Americans are less likely than whites to have access to private pensions and more likely to rely on government retirement programs, they may leave the labor force later in life or substitute part-time work for full-time work until they can take advantage of Social Security benefits.

Net of the effects of health status, discrimination, work history, and socioeconomic status, we suggest that another reason for expecting race differences centers on the nonmonetary qualities of jobs. In other words, the poor intrinsic value of many jobs available to African Americans may motivate them to seek alternatives to full or continuous participation. Low levels of

intrinsic value in available jobs may influence some African American men to report disability as a status rather than retirement, to reduce their commitment to the labor force by reducing their hours, or to leave the labor force and not reenter. This may be even more likely when Social Security or disability payments provide a relatively fair substitute for the economic rewards available in the job market — compensation which is on average low for African Americans.

Sociological literature contains support for this idea. For example, Jencks, Perman and Rainwater (1988) divide the desirability of jobs into two categories: monetary and nonmonetary. They find that when African Americans and whites rate their jobs, only one-fifth of the difference in their ratings is attributable to monetary concerns. The remaining difference in job ratings has to do with the nonmonetary characteristics (e.g., level of supervision, risk of job loss, repetition in job tasks). Jencks, Perman and Rainwater (1988) show that, in general, jobs held by African Americans are less desirable than those held by whites.

To summarize, we have argued that African Americans and whites differ in terms of the central economic, demographic, and social factors that influence labor force behavior in later life. In addition, we believe that African Americans and whites with the same levels of these important attributes respond differently when faced with decisions about the level of labor force attachment they maintain in the later stages of the working life cycle. The model we present here assumes that individuals consider both economic and noneconomic factors, as well as perceived labor market alternatives, when evaluating their current employment or nonemployment. Our model is consistent with life-course perspectives on work and retirement put forth in the sociological and economic literatures, but we do not presume to capture all the relevant motivating factors.[3] Instead, we hope to establish a foundation for more extensive studies of the late-life labor force behavior of nonwhites and women. We start by providing descriptive information about variation in labor force transitions by race and by developing insights into the employment-related behavioral differences among African American and white males. Below we discuss the methods that allow us to capture a view of this complex and dynamic behavior.

Data and Methods

We use data from a pooled sample of respondents from the 1984 and 1985 panels of the Survey of Income and Program Participation (SIPP). Pooling the data allows us to obtain a larger sample. The SIPP is designed to provide longitudinal data for a nationally representative sample of the civilian, noninstitutionalized population (see Kasprzyk et al. 1987 for a detailed discussion of the SIPP). Panels are fielded for two and a half years; the 1984 panel began in October 1983, and the 1985 panel began in February 1985. Because of the staggered interviewing, our data covers the period from January 1984 to December 1986 (24 months of observation).

This study includes non-Hispanic African American and non-Hispanic white men aged 55 to 69 who were ether included in both the first and seventh wave of each survey, or who died or became institutionalized during this period.

Those who left the sample for other reasons (e.g., those who refused to continue participating, and those who moved out of the interview area) are excluded from the sample, resulting in a 15.7% reduction. We account for features of the sample design and patterns of attrition by applying modified sample weights. We begin with the cross-sectional weight provided by the Census Bureau and included in the SIPP data for the initial wave of observation. We then adjust these weights to accommodate attrition over the period of study, taking into account differential attrition by race and panel (1984/1985). Our strategies for weighting the sample data were guided by discussions in Kasprzyk et al. 1989. Details of this weight adjustment process are available from the authors.

Indicators traditionally used to examine complex labor force behavior are somewhat constrained, often based on cross-sectional or retrospective summary indicators of labor force participation or nonparticipation (e.g., Parsons 1980). In addition, measures of subjective retirement can be clouded by other mitigating options, such as subjective disability (Gibson 1987). Finally, some researchers have employed receipt of pension income as a dimension of labor force status. This strategy may leave out many disadvantaged persons because of their lifelong poor market positions (Jackson & Gibson 1985).

We overcome these limitations by developing a cross-sectional measure of different levels of labor force participation (e.g., part-time vs. full-time work). We can then generate this measure for different time points, which allows us to infer changes in status (e.g., reentry, full withdrawal).

Our labor force measure is based primarily on a summary of work experiences in the month before each interview. We include four labor force statuses: (1) full-time work, which includes individuals who usually worked 35 or more hours per week; (2) part-time work, which includes individuals who usually worked fewer than 35 hours per week; (3) unemployed, which includes individuals who experienced spells of unemployment during the month; and (4) not in the labor force, which includes individuals who had no work attachment in the month prior to interview.[4] For our transition analysis, we added another category, which includes those individuals who either died or were institutionalized during the observation period. See Chirikos & Nestel 1989 for a discussion of the importance of the mortality outcome in studies of late-life labor force behavior: data on reason for attrition were provided to the authors by the U.S. Bureau of the Census.

The independent variables in the multivariate analyses are taken from wave 1 of each panel. Disability status is a central variable in the study of late-life labor force behavior (for a recent discussion of current research in this area, see Mutchler et al. 1994). We measure disability status by response to the question "Does _____ have a physical, mental or other health condition that limits the kind or amount of work _____ can do?" This question is asked of the respondent in wave 1 (for those younger than 70) and is a standard question for the presence of a work-limiting health condition (1 = presence of a work-limiting health condition).[5]

We include a dummy variable for marital status because late-life labor force behavior is known to vary by this characteristic (1 = married). We also control for age of respondent inasmuch as our study sample crosses younger age respondents who are less likely to be out of the labor force and respondents

from older age groups who are more likely to be out of the labor force due both to normative pressures and institutional factors (e.g., eligibility for public and private pensions and Medicare). To capture the anticipated nonlinearity in the age effect associated with program eligibility requirements, we also include in our models an age spline function. Persons aged 55 to 61 are assigned a zero and persons aged 62 and older are assigned a value equivalent to their age minus 61 (thus, 62 = 1, 63 = 2, 64 = 3, and so on). The age spline function is used in conjunction with the age variable in the multivariate models.

Current income is measured as the respondent's reported total personal income in the four months prior to first interview (own income only; we use the natural log of income in the multivariate analyses). Number of years of completed education for each respondent is included as an indicator of social class. Finally, we include a dummy variable to indicate if the respondent was part of the 1984 or 1985 SIPP panel (1 = 1984 panel). This variable is designed to pick up any gross differences in sample design and period effects across the two panels. Descriptive information based on time 2 (wave 7) labor force status and the variables outlined above are presented in the Appendix. We leave it to the reader to inspect these data.

We present the results of our analysis in two phases. First, we present transition rate matrices for both races combined and for each race group, capturing the stability and change in work status (and attrition) between wave 1 and wave 7. Second, we estimate multivariate logistic regression models of labor force transitions. Separate analyses are conducted for each race group. Because the dependent variable is categorical and unordered, we use multinomial logistic regression (e.g., Hanushek & Jackson 1977; Maddala 1983). More details on the multivariate analysis are given below.

Results

Table 1 describes the labor force transition rates for the 24-month period for both races combined. The table shows that a great deal of variability occurs in terms of how much movement surrounds a given labor force category. For example, of the men reporting they were not in the labor force at first observation, nearly 86% were not in the labor force at the final observation. This suggests substantial stability among nonworkers in this age group.[6] Of those who did leave the not-working status, 8.0% died or became institutionalized, and 5.5% reentered the labor force either as full-time or part-time workers.

Full-time workers were a less stable group: Just under 69% reported this status at final observation. In contrast, fewer than half (47%) of those reporting part-time work at first observation reported the same status at final observation. Nearly 20% had moved to full-time work 24 months later, and 30% left the labor force altogether. Not surprisingly, the unemployed were a very unstable group: Fewer than 14% continued to report this status at final observation. The largest share of those unemployed at first observation left the labor force during the two-year period (46.5%).

Among the four labor force statuses identified in wave 1, Table 1 shows unique patterns of mortality and health by labor force status at time 2 (wave 7).

TABLE 1: Labor Force Transitions for Full Sample, Wave 1 to Wave 7[a]

	Labor Force Status (Wave 7)					
Labor force status (wave 1)	Full-Time (%)	Part-Time (%)	Unemployed (%)	Not Working (%)	Died/ Inst. (%)	Total (Wave 1) (%)
Full-time	68.9	8.1	2.4	18.5	2.0	100.0 (1,684)
Part-time	19.4	47.2	.2	29.7	3.4	100.0 (302)
Unemployed	22.4	12.3	13.9	46.5	5.0	100.0 (107)
Not working	2.0	3.5	.7	85.8	8.0	100.0 (1,640)
Total (wave 7)	(1,281)	(348)	(63)	(1,860)	(181)	

[a] Total unweighted counts for each labor force status in wave 1 and wave 7 are given in parentheses (weighted sample size = 5,249, unweighted sample size = 3,733). Data are weighted as discussed in text. Differences between sum of categories and reported totals due to rounding error.

Full-time workers are the least likely to die or become institutionalized, and those who are not working are the most likely to have died or become institutionalized. These relationships are most certainly confounded with age differences across the four labor force statuses identified here. Some of the age effect will be captured in our multivariate models (discussed below).

Next, we turn to a discussion of similar transition rate matrices broken down by race — the primary focus of our study (see Table 2). The reader should be aware that some cells in the matrices contain sampling zeros. Table 2 shows the transition rates for the white male and African American male samples.

While there are some notable similarities between older African American men and older white men in labor force transitions, there are some interesting differences as well.[7] One similarity is the relative stability of the not-working status for each group across the two time points (86.3% for whites and 81% for African Americans). Another is the instability of part-time work status. Part-time work status was the most unstable of the two working statuses reported here (full-time and part-time), which supports the findings of other researchers (see Hayward, Hardy & Liu 1994). Part-time work is a function of many factors at this age, including (1) a preference for partial retirement over full retirement (see Honig & Hanoch 1985), (2) the need to cover income

TABLE 2: Labor Force Transitions by Race, Wave 1 to Wave 7

	White Sample					
	Labor Force Status (Wave 7)					
Labor force status (wave 1)	Full-Time (%)	Part-Time (%)	Un-employed (%)	Not Working (%)	Died/ Inst. (%)	Total (Wave 1) (%)
Full-time	68.6	8.1	2.6	18.6	2.2	100.0 (1,563)
Part-time	18.8	47.8	0.0	29.7	3.7	100.0 (281)
Unemployed	22.2	11.0	15.9	46.4	4.5	100.0 (94)
Not working	2.1	3.6	.7	86.3	7.4	100.0 (1,485)
Total (wave 7)	(1,178)	(326)	(60)	(1,702)	(157)	

	African American Sample					
	Labor Force Status (Wave 7)					
Labor force status (wave 1)	Full-Time (%)	Part-Time (%)	Un-employed (%)	Not Working (%)	Died/ Inst. (%)	Total (Wave 1) (%)
Full-time	73.9	8.0	0.0	17.4	.7	100.0 (121)
Part-time	28.1	38.7	3.4	29.8	0.0	100.0 (21)
Unemployed	24.3	20.5	0.0	46.8	8.4	100.0 (13)
Not working	1.3	2.2	1.1	81.0	14.3	100.0 (155)
Total (wave 7)	(103)	(22)	(3)	(158)	(24)	

[a] Total unweighted counts for each labor force status in wave 1 and wave 7 are given in parentheses. Data are weighted as discussed in the text (weighted African American sample size = 435, unweighted African American sample size = 310 and weighted white sample size = 4,796 and unweighted white sample size = 3,423). Differences between sum of categories and reported totals are due to rounding error.

shortfalls in the context of what would otherwise be a time of full retirement, and (3) an inability to find full-time employment in a slack job market for the reasons noted earlier in the article.

We also found two major differences between the two races. First, among those who started the survey in the labor force (either full-time or part-time), the fraction who died or became institutionalized (mostly in nursing homes at this age) is much lower for African American men than for white men. Yet among those who began the observation period not working, twice as many African American men died or were institutionalized compared to white men (14.3% vs. 7.4%). This implies that health may be a more important predictor of labor force status and transitions for African American men than it is for white men.

A second major difference by race centers on part-time work status. African American men who begin the observation period as part-time workers are much more likely to move into full-time employment than are their white counterparts (28.1% vs. 18.8%). This difference in the rate of transition into full-time work is likely indicative of the life-long instability in labor force attachment (and opportunity) among African American males.

To establish the statistical significance of the differences identified in Table 2, we fit a series of hierarchical loglinear models to the origin by destination by race cross-classification. The fit of the various models is shown in Table 3. When we compare the results from models 2 and 1, we can see that the majority of the association in the table is between the origin and destination status, but the model does not adequately fit the data. Including interactions for race-by-origin (model 3) or race-by-destination (model 4) or both interaction terms simultaneously (model 5) improves the fit of the data, but the overall fit of the model is still inadequate. Thus, the saturated model, including the three-way interaction between origin, destination, and race, is required to fit the observed association. This suggests that the patterns shown in Table 2 are significantly different for this sample of white and African American men.

Multivariate Analyses

As indicated by the results presented in Table 3, the overall patterns of labor force status and transitions observed in Table 2 represent statistically significant race differences. At this point we turn to the possible effects of background characteristics on these transitions.

For this phase, we estimate separate multinomial logistic regression models for the African American and white samples, using weighted data, as discussed earlier. This form of the multinomial logit model provides a comparison of categorical outcomes through a series of binomial contrasts. We employ the CATMOD procedure from the SAS statistical software to estimate these models. More formally, the model for this analysis is given below:

$$Log(P_j/P_m) = Xb_j$$

This model specifies the log-likelihood of being in labor force category j relative to labor force category *m* (or sample exit through death or institutionalization)

TABLE 3: Fit Statistics for Alternate Loglinear Models Assessing the Association between Wave 1 Labor Force Status (A), Wave 7 Labor Force Status (B), and Race (C)[a]

Marginals fit to the data	Likelihood Ratio Chi-Square	df	p	D
1. A B C	3,738.991	31	.000	.368
2. A * B C	56.821	19	.005	.016
3. A * B A * C	42.287	16	.000	.012
4. A * B B * C	43.339	15	.000	.011
5. A * B A * C B * C	30.951	12	.002	.007

[a] Data are weighted as discussed in text.

as a function of X, the set of explanatory variables given earlier, and b_j, the vector of parameters associated with category j.

Our multivariate logistic regression analyses focus on labor force status at time 2 (wave 7 of the SIPP observation period) as the outcome variable. The predictors in this model include the variables described above in the Data and Methods section. We also include in these models respondent's labor force status at origin (wave 1), and exclude from the model a reference group for one of the labor force statuses at time 1 that is the same as the reference category at time 2 (e.g., full-time origin is not included when the reference category at time 2 is full-time work). This allows us to evaluate more intuitively the impact of change in labor force status across the two time points. For example, when we compare working full-time with working part-time at time 2, we generate coefficients for start statuses of not working and part-time working at time 1. Here we are able to estimate the log-likelihood of moving from not-working status at time 1 to either full-time or part-time status at time 2 (reentry into the labor market).

In general, the coefficients represent the estimated log-likelihood of moving to one of two labor force statuses or to a single labor force status or sample exit by dying or being institutionalized at wave 7. In the first part of the analysis reported in Table 4, the reference category in contrasts 1 through 3 (full-time vs. part-time, full-time vs. not working, full-time vs. death/institutionalization) is working full-time at time 1. Normally, these are the only results that would be reported. However, to give the reader a full picture of all possible transition contrasts, we present in contrasts 4 through 6 the nonredundant labor force comparisons not captured in the first three contrasts (not working vs. part-time, not working vs. death or institutionalization, and part-time vs. death or institutionalization).[8]

TABLE 4: Results from Logistic Regression Models of Transitions in Labor Force Status among Older African American and White Men[a]

	Labor Force Status (Wave 7)					
	Part-time vs. Full-time		Not Working vs. Full-time		Death/Inst. vs. Full-time	
	African American	White	African American	White	African American	White
Age (years)	.616** (.185)	.235** (.038)	.442** (.121)	.357** (.030)	.608** (.184)	.259** (.054)
Age (spline)	-.746** (.276)	-.032 (.063)	-.547** (.210)	-.224** (.052)	-.694* (.286)	-.093 (.082)
Marital status (1=married)	.021 (.558)	-.145 (.173)	-.565 (.418)	-.207 (.139)	-1.084+ (.560)	-.460* (.204)
Work-limiting hlth. cond. (1=yes)	2.780** (.676)	.300+ (.168)	1.933** (.562)	.553** (.135)	2.779** (.729)	1.454** (.191)
Income (logged)	-.082 (.133)	.0003 (.048)	.364** (.137)	.123** (.041)	.290 (.251)	-.051 (.054)
Education (years)	-.002 (.059)	.001 (.019)	-.050 (.045)	-.072** (.016)	-.024 (.065)	-.038 (.025)
Labor force status (wave 1)						
Full-time	—	—	—	—	—	—
Part-time (1=yes)	2.630** (.687)	2.610** (.170)	1.273* (.603)	1.311* (.173)	—[b]	1.194** (.335)
Not working (1=yes)	1.831+ (.970)	2.049** (.227)	5.586** (.832)	4.401** (.180)	6.530** (1.276)	3.654** (.251)
Panel flag (1=1984)	-.964+ (.523)	-.152 (.125)	.626 (.396)	-.110 (.100)	.591 (.537)	.040 (.163)
Intercept	-38.084** (11.100)	-16.005** (2.255)	-30.162** (7.145)	-22.525** (1.808)	-42.913** (11.105)	-17.841** (3.175)

[a] See text for definitions of variables. Log-likelihood coefficients with standard errors in parentheses. The fit of the model to the data for the African American sample is $\chi^2 = 427.0$ (df = 30) and the fit of the model to the data for the white sample is $\chi^2 = 3{,}744.6$ (df = 30). The data are weighted as discussed in the text. Boxed pairs of coefficients represent statistically significant differences across the African American and white samples.

TABLE 4 Results from Logistic Regression Models of Transitions in Labor Force Status among Older African American and White Men[a] (Continued)

	Labor Force Status (Wave 7)					
	Part-time vs. Not Working		Death/Inst. vs. Not Working		Death/Inst. vs. Part-time	
	African American	White	African American	White	African American	White
Age (years)	.174 (.186)	-.122** (.041)	.166 (.144)	-.098+ (.051)	-.008 (.230)	.024 (.060)
Age (spline)	-.199 (.270)	.191** (.060)	-.147 (.203)	.131+ (.072)	.052 (.330)	-.061 (.087)
Marital status (1=married)	.586 (.546)	.062 (.164)	-.519 (.390)	-.252 (.171)	-1.052+ (.655)	-.314 (.222)
Work-limiting hlth. cond. (1=yes)	.867 (.601)	-.253+ (.147)	.847+ (.491)	.901** (.158)	-.020 (.757)	1.154** (.202)
Income (logged)	-.446** (.154)	-.122* (.049)	-.075 (.229)	-.173** (.047)	.371 (.259)	-.051 (.061)
Education (years)	.048 (.060)	.073** (.019)	.026 (.049)	.034 (.021)	-.022 (.076)	-.039 (.027)
Labor force status (wave 1)						
Full-time (1=yes)	3.755** (.739)	2.352** (.170)	-.944 (1.017)	.747** (.201)	—[b]	1.416** (.329)
Part-time (1=yes)	5.113** (.861)	3.650** (.171)	—[b]	.629* (.295)	—	—
Not working (1=yes)	—	—	—	—	13.183** (1.207)	3.021* (.312)
Panel flag (1=1984)	-1.519** (.529)	-.042 (.121)	-.035 (.379)	.150 (.144)	1.555* (.639)	.192 (.177)
Intercept	-11.676 (11.136)	4.169+ (2.432)	-11.807 (8.670)	3.938 (3.007)	-13.315 (13.992)	-3.253 (3.578)

[b] Not enough cases for computation of effects.

* p < .05 ** p < .01 + .05 < p < .10 (two-tailed test)

For each labor force status contrast, the results from the African American model are presented in the first column and the results from the white model are presented in the second column to facilitate comparisons. Differences in racial contrasts for each pair of coefficients significant at the .05 probability level or better are enclosed in boxes. Comparative results not enclosed in boxes suggest that observed race-specific patterns are not significantly different across the African American and white samples.

The results in Table 4 show strong positive effects of initial labor force status on the likelihood of changing labor force status for each race group. In general, this suggests that net of the other variables in the model, there is considerable labor force stability among most categories of work. This supports the interpretation of some of the results from Tables 1 and 2. With two exceptions, these results also show that the association between starting and ending work status is essentially the same for African Americans and whites, *net of health, socioeconomic status, and demographic characteristics.*

The two exceptions to this pattern occur among those who start the observation period in the not-working status. Both African American and white men in this group are more likely to die or become institutionalized than those who start the period working either part-time or full-time. However, among men who start the period not working, African Americans are more likely than whites in these two contrasts to die or become institutionalized. This suggests that African American men who are not working at the start of the observation period have higher chances of dying over the two-year period than white men.

The effects of age for the two groups differ only in the part-time versus full-time contrast — the decision of how many hours to work. The greater likelihood for older workers to report other than *full-time work* as an ending status is more pronounced for African American men than for white men. The age spline effect (capturing the flattening of the age effect beginning at age 62) is also more substantial for African American men than for their white counterparts when we contrast death/institutionalization and full-time work.

Together these effects demonstrate that African American men are less likely to work full-time at an earlier point in the working life cycle than white men. The effects of health and socioeconomic status also differ between the groups. African American men with work-limiting health conditions are significantly more likely than similar whites to choose part-time instead of full-time work, or to leave the labor force altogether. Higher income men of both races are significantly more likely than their low-income counterparts to be nonworkers rather than part-time workers, yet the difference is significantly larger among African American men than among white men. Thus African American males with sufficient income are more likely to take advantage of this resource to leave the labor force at the later stages of the working life cycle, net of the other variables in the model.

Taken as a whole, these results suggest that although the transition patterns show more similarities than differences among African American and white men, there are some meaningful differences in the *process* associated with those transitions. This appears to be particularly relevant to the decision to work full-time or part-time rather than leave the labor force altogether. In particular, older African American men apparently respond more strongly to health and socioeconomic status than whites and leave full-time work at an earlier age.

Although we are not able to test the "good jobs" hypothesis directly with the SIPP data, racial differences in attachment to the labor force are likely at least in part

to be the result of older African American men rejecting work because of the desirability of the jobs available to them. In other words, our results show that African American and white men respond differently to decisions about work given similar levels of income, age and health. This differential response may be partly related to the kinds of work done, physical stress of the job, and the job's intrinsic reward structure. The results of this analysis for this hypothesis are only suggestive and should be followed up with research that more explicitly tests these ideas.

Discussion

As we have discussed, permanent withdrawal from the labor force in later life is for many persons a process and not a single event. This process often includes spells of unemployment before an involuntary and complete exit from working life. The process may also include either partial retirement or part-time work. Further, many older persons leave their major lifetime occupations only to return to the labor force within a short period of time.

With these issues in mind, this study is designed to assess differences among older white and African American men in their work and nonwork characteristics and in their transition behavior over a 24-month observation period. Existing research documents important differences in labor force participation *levels* between African American and white men measured at single points in time throughout the adult life cycle. In this study, we document meaningful differences in labor force *transition rates* using panel data at later stages of the working life cycle. Multivariate analyses of labor force transitions indicate that statistical associations among key labor force statuses and a set of theoretically important predictors, such as health and socioeconomic status, differ significantly by race. Accordingly, we conclude that racial differences in later life labor force outcomes are important and should continue to be explored.

For example, as we noted earlier, some observers claim that African Americans are more likely to assume the disabled sick role in old age than whites, which may partially account for their greater proportion of the "unretired retired" (Gibson 1987; Jackson & Gibson 1985). One of the assumptions is that subjectively assessing oneself as disabled may be a way of coping with the inability to find a desirable job or to retire in a more traditional fashion.

However, evidence provided here and elsewhere (Hayward, Friedman & Chen 1996) suggests that the African American "unretired retired" are not simply claiming poor health as a social psychological adjustment mechanism. As we found, African American men who began the observation period not working were twice as likely as white men to die or become institutionalized, indicating that they have objectively poorer health. This may show that the relationship between race, health, and labor force participation is more complex, requiring further evaluation. That is, death and institutionalization are "objective" states whereas claims of disability reflect a more nebulous mix of subjective and objective realities. (Note that African Americans are on average much less likely to make use of institutions in old age than whites; see Burr 1990.) Given differences in the transition to the death/institutionalized state in the SIPP data, African Americans who are not working in old age would appear to be in objectively worse health than their white counterparts and this poor health negatively impacts their labor force participation. For at least some substantial portion of older African American men, they are not simply claiming poor health as

a social psychological adjustment mechanism. Rather they are in worse health and this effects their level of labor force participation.

Another potentially fruitful avenue for future research is suggested by the hypothesis that older African American men, who have suffered economic disadvantage throughout their working lives, may be prone to a less stable attachment to the labor force if there are few or no desirable jobs available. In such cases, they likely prefer to reduce their work hours or remain out of the labor force if other sources of income (e.g., disability income, Social Security) come close to meeting their economic needs. While our work does not directly assess this possibility due to data limitations, the results suggest these explanations may have some utility.

What does the future hold for the labor force behavior of older African American and white men? During the last 10 years, American corporations have engaged in downsizing through massive layoffs, often laying off persons in high skilled, middle and upper white-collar positions as well as persons in less skilled positions. A disproportionate number of these jobs are held by white men. For many older white men, these layoffs have disrupted otherwise orderly careers and created difficult economic circumstances along with uncertain retirement expectations and horizons. If this restructuring continues, the labor force behavior of older white men may begin to resemble the less stable behavior characteristic of large segments of the older African American male population. This is another topic for future research.

Finally, our study also contributes to the continuity of research into labor force behavior among older persons in the U.S. The behavior of earlier cohorts of men has been thoroughly chronicled in national surveys such as the Retirement History Survey, the National Longitudinal Surveys of Older Men and Women, and the Panel Study of Income Dynamics. And the long anticipated Health and Retirement Survey will allow researchers to examine the behavior of a later set of birth cohorts not captured elsewhere. The present study and others like it that rely on the SIPP allow researchers to fill the gap between the studies just mentioned. Such descriptive research allows for continuity in the coverage of birth cohorts and provides the opportunity for establishing policy and theory based on a continuum of information.

Notes

1. Parsons (1980) uses a somewhat dubious control for health in his cross-sectional models: subsequent mortality experience after the observation point. One of the contributions of the present research is a longitudinal evaluation of multifaceted labor force behavior with health measured at time 1 and mortality, if present, measured at time 2.

2. Research on the association between health status and labor force behavior at the end of the working life cycle for the general population is well developed. Many studies have demonstrated a negative association between health and labor force activity (e.g., Ekerdt, Bosse & LoCastro 1983; Hardy 1982; Hardy & Pavalko 1986; Hausman & Wise 1985; Hayward & Grady 1990; Hayward, Friedman & Chen 1996). While health limitations typically affect work status, they do not necessarily signal permanent labor force withdrawal. Instead, they may result in reduced participation, a job change, or temporary withdrawal (Hausman & Wise 1985; Hayward & Grady 1990). In addition, persons may recover from poor health or temporary disabilities and reenter the labor force (see Hayward, Friedman & Chen 1996). We expect that the linkage between race, health, and the dynamics of labor force behavior are based at least partially on the actual health differences by race.

3. One variable not included in the present analysis of race differences in labor force transitions is pension eligibility. Because of the location of the relevant questions in the SIPP protocol, no information is available that permits an evaluation of labor force behavior when using a 24-month observation period (the necessary questions are simply not asked at appropriate

APPENDIX: Descriptive Characteristics of African American and White Men[a]

	African Americans	Whites
Labor force status (percent) (wave 7)		
Full-time	32.7	35.0
Part-time	6.8	9.6
Not working	52.4	50.7
Died/institutionalized	8.0	4.7
Age (mean years)	61.5 (4.3)	61.7 (2.7)
Married (percent)	63.2	82.7
Work-limiting health condition (percent)	42.5	29.4
Income (median $)	3,000	5,585
Education (mean years)	8.6 (4.3)	11.8 (3.5)
Total (unweighted) sample size	290	3,265

[a] See text for definition of variables. Standard deviations in parentheses.

points in the survey.) Although ideally we would like to have such a measure in our model, we believe this is not a serious problem given our goals. Further, we have estimated similar models of labor force transition for these two race groups based on an eight-month observation period, which does allow for the estimation of pension eligibility. Results of a series of models that include a pension measure show that it has no effect on the other variables, regardless of whether it is in the model or not. That is, the absence of a measure of pension eligibility from the model does not change the statistical effects or the interpretation of other measures in the model. Second, measures that are included in the present analysis (e.g., age, education, and income) proxy some of the expected effects of pension eligibility.

4. We recognize that considerable heterogeneity within each labor force category may exist. For example, the not-working category includes those who are voluntarily and involuntarily not working. Similarly, the part-time worker category likely contains persons who have withdrawn from their lifetime occupations and continue to work for their own satisfaction, in addition to persons who must work to supplement their retirement income. Consequently, caution should be exercised when interpreting behavior within categories.

5. A few additional persons are categorized as disabled in subsequent waves of the sample. In the SIPP, once respondents are classified as disabled, they are not asked again about this issue. Thus, because of the design of the SIPP, it is not possible to determine whether a person recovers from such a condition. Note also that there have been several attempts to determine if self-evaluation of health or disability status differs by race (e.g., Angel & Thoits 1987; Gibson

1991b; Mutchler & Burr 1991). Although there are theoretical reasons to expect differences, research is not conclusive.

6. Results of wave-to-wave analyses of labor force status for a similar sample of men from the SIPP (not reported here) suggest that stability in labor force behavior over the two-year observation period is characteristic of about two-thirds of the sample. More than 40% of the men were classified as not in the labor force at all seven observation points (four-month reference periods), while about one-fifth of the men worked full-time for the entire period. An additional 6% of the men experienced transitions between full-time and part-time work only, neither leaving nor entering the labor force, but rather moving between full-time and part-time work. About 8% left the labor force, and about 1% entered it. About 15% experienced two or more changes in status other than changes in hours: These include retirements followed by reentries, reentries followed by reexits, or other multiple transitions. Taken as a whole, these data indicate substantial movement in levels of labor force participation during these two years. This is similar to findings reported by other researchers (see Anderson et al. 1984).

7. Researchers have noted that older African American and white men observed at a single point in time display different patterns of labor force participation, and these race groups exhibit substantial variation in expected number of years of working life (see Smith 1986; Welch 1990). An analysis of labor force status by race at wave 1 shows that African American men are less likely than similar white men to be in the labor force (results not reported here). These cross-sectional racial differences are statistically significant and comparable to findings from other data sources (U.S. Bureau of Labor Statistics 1994).

8. The coefficients for the contrasts in columns 4-6 can be generated by algebraic manipulation. However, this method of presentation makes comparisons easier.

References

Anderson, Kathryn H., Richard V. Burkhauser and J.S. Butler. 1984. "Work after Retirement: A Hazard Model of Labor Market Reentry." Institute for Research on Poverty, Discussion Paper no. 768-84.

Angel, Ronald J., and Peggy A. Thoits. 1987. "The Impact of Culture on the Cognitive Structure of Illness." *Culture, Medicine, and Psychiatry* 11:465-94.

Atchley, Robert C. 1982. "Retirement as a Social Institution." *Annual Review of Sociology* 8:263-87.

Beck, Scott H. 1985. "Determinants of Labor Force Activity among Retired Men." *Research on Aging* 7:251-80.

Burr, Jeffrey A. 1990. "Race/Sex Comparisons of Elderly Living Arrangements: Factors Influencing the Institutionalization of the Unmarried." *Research on Aging* 12:507-30.

Chirikos, Thomas N., and Gilbert Nestel. 1985. "Longitudinal Analyses of Functional Disabilities among Older Men." *Journal of Gerontology* 40:426-33.

____. 1989. "Occupation, Impaired Health, and the Functional Capacity of Men to Continue Working." *Research on Aging* 11:174-205.

DeViney, Stanley, and Angela M. O'Rand. 1988. "Gender-Cohort Succession and Retirement among Older Men and Women, 1951 to 1984." *The Sociological Quarterly* 29:529-40.

Ekerdt, David J., Raymond Bosse, and Joseph S. LoCastro. 1983. "Claims That Retirement Improves Health." *Journal of Gerontology* 38:231-36.

Gibson, Rose C. 1987. "Reconceptualizing Retirement for Black Americans." *The Gerontologist* 27:691-8.

____. 1991a. "The Subjective Retirement of Black Americans." *Journal of Gerontology* 46:S204-209.

____. 1991b. "Race and the Self-Reported Health of Elderly Persons." *Journal of Gerontology* 46:S235-42.

Gohmann, Stephen F. 1990. "Retirement Differences among the Respondents to the Retirement History Survey." *Journal of Gerontology: Social Sciences* 45:S120-27.

Gustman, Alan L. and Thomas L. Steinmeier. 1986. "A Disaggregated, Structural Analysis of Retirement by Race, Difficulty of Work and Health." *The Review of Economics and Statistics* 68:509-13.

Hanushek, Eric A., and John E. Jackson. 1977. *Statistical Methods for Social Scientists*. Academic Press.

Hardy, Melissa A. 1982. "Social Policy and Determinants of Retirement: A Longitudinal Analysis of Older White Males, 1969-75." *Social Forces* 60:1103-22.

Hardy, Melissa A., and Eliza K. Pavalko. 1986. "The Internal Structure of Self-Reported Health Measures among Older Male Workers and Retirees." *Journal of Health and Social Behavior* 27:346-57.

Hausman, Jerry A., and David A. Wise. 1985. "Social Security, Health Status and Retirement." Pp. 159-81 in *Pensions, Labor, and Individual Choice*, edited by David A. Wise. University of Chicago Press.

Hayward, Mark D., and William R. Grady. 1990. "Work and Retirement among a Cohort of Older Men in the United States, 1966-1983." *Demography* 27:337-56.

Hayward, Mark D., Melissa A. Hardy, and Mei-Chen Liu. 1994. "Work after Retirement: The Experiences of Older Men in the United States." *Social Science Research* 23:82-107.

Hayward, Mark D., Samantha Friedman, and Hsinmu Chen. 1996. "Race Inequities in Men's Retirement." *Journal of Gerontology: Social Sciences* 51B:S1-S10.

Henretta, John C. 1992. "Uniformity and Diversity: Life Course Institutionalization and Late-Life Work Exit." *The Sociological Quarterly* 33:265-79.

Henretta, John C., and Angela M. O'Rand. 1983. "Joint Retirement in the Dual Career Family." *Social Forces* 64:1032-45.

Honig, Marjorie, and Giora Hanoch. 1985. "Partial Retirement as a Separate Model of Retirement Behavior." *Journal of Human Resources* 10:21-46.

Jackson, James S., and Rose C. Gibson. 1985. "Work and Retirement among the Black Elderly." Pp. 193-222 in *Current Perspectives on Aging and the Life Cycle: A Research Annual*. Vol. 1, *Work, Retirement, and Social Policy*, edited by Zena S. Blau. JAI Press.

Jencks, Christopher, Lauri Perman, and Lee Rainwater. 1988. "What is a Good Job? A New Measure of Labor-Market Success." *American Journal of Sociology* 93:1322-57.

Kasprzyk, Daniel, Patricia Doyle, Arthur Goldstein, and David B. McMillen. 1987. *Survey of Income and Program Participation User's Guide*. U.S. Bureau of the Census.

Kasprzyk, Daniel, Greg J. Duncan, Graham Kalton, and M.P. Singh. 1989. *Panel Surveys*. Wiley.

Kotlikoff, Laurence J., and David A. Wise. 1989. *The Wage Carrot and the Pension Stick*. Upjohn Institute.

Lazcko, Frank. 1989. "Early Exit and the Employment of Older Workers in the 1990s." Paper presented at the XIV International Congress of Gerontology. Acapulco, Mexico.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.

Mutchler, Jan E., and Jeffrey A. Burr. 1991. "Racial Differences in Health and Health Care Service Utilization: The Effect of Socioeconomic Status." *The Journal of Health and Social Behavior* 32:342-56.

Mutchler, Jan E., Jeffrey A. Burr, Michael P. Massagli, and Amy M. Pienta. 1994. "Health and Labor Force Status in Later Life." Paper presented at the annual meeting of the Population Association of America, Miami.

Parsons, Donald O. 1980. "Racial Trends in Male Labor Force Participation." *American Economic Review* 70:911-20.

Parnes, Herbert S., and Gilbert Nestel. 1981. "The Retirement Experience." Pp. 155-97 in *Work and Retirement: A National Longitudinal Study of Men,* edited by Herbert S. Parnes. MIT Press.

Pienta, Amy M., Jeffrey A. Burr, and Jan E. Mutchler. 1994. "Women's Labor Force Participation in Later Life: The Effects of Early Work and Family Experiences." *Journal of Gerontology* 49:231-9.

Quinn, Joseph F., Richard V. Burkhauser, and Daniel A. Myers. 1990. *Passing the Torch: The Influence of Economic Incentives on Work and Retirement.* Upjohn Institute.

Smith, Shirley J. 1986. "Worklife Estimates: Effects of Race and Education." Bulletin no. 2254. Bureau of Labor Statistics.

Szinovacz, Maximiliane. 1985. "Women and Retirement." Pp. 327-53 in *Growing Old in America,* edited by Beth Hess and Elizabeth Markson. Transaction Books.

U.S. Bureau of Labor Statistics. 1994. "Employment and Earnings." 42:1. U.S. Department of Labor: USEPO.

Welch, Finis. 1990. "The Employment of Black Men." *Journal of Labor Economics* 8:S26-74.

Track Mobility in Secondary School

MAUREEN T. HALLINAN, *University of Notre Dame*

Abstract

Tracking, a pervasive organizational practice in American secondary schools, creates instructional groups that are homogeneous with respect to student ability. It is a highly controversial practice frequently criticized because it locks students into a particular track level for an entire school year or longer. This study examines several theoretical ideas about the basis of track selection and about track mobility patterns. Relying on longitudinal data from over 2,000 students, the research examines the frequency of track mobility and the direction of change in track assignments over students' four year high school careers. The data show evidence of considerably more mobility than is typically assumed. The findings relate mobility to student and school characteristics and to the curriculum. The results demonstrate how track mobility can provide differential learning opportunities to pupils, particularly females and low-income students.

Tracking, or the grouping of students by ability for instruction, occurs in most American secondary schools in both English and mathematics. A typical track structure in English consists of four levels: advanced, honors, regular, and basic. The track structure in mathematics tends to be similar, but may include a fifth level, often designated as very basic. Larger high schools may have a more elaborate hierarchy of track levels, while smaller schools may collapse some track categories.

The practice of tracking students for instruction is highly controversial. Advocates of tracking argue that the practice facilitates student learning by enabling teachers to gear their lessons to their students' ability levels. Critics claim that tracking disadvantages low-track students because the quality of instruction in the low tracks is inferior to that provided at higher level tracks and because it damages students' self esteem. Moreover, critics claim that track

* *The author gratefully acknowledges support from the National Science Foundation, Grant #SES9009298, and the U.S. Department of Education, Office of Educational Research and Improvement, Grant #R117E10139. I also wish to thank Warren Kubitschek for assistance with data collection, data preparation and statistical analysis, and Sylvia Phillips and Ann Power for manuscript preparation. I am indebted to Karl Alexander and Adam Gamoran for several very helpful comments on the paper and insightful suggestions that strengthened the argument and the analysis. Finally I am grateful to the principals, counselors, teachers and students who provided the data for the research. Direct correspondence to Maureen T. Hallinan, Department of Sociology, University of Notre Dame, Notre Dame, IN 46556.*

 Social Forces, March 1996, 74(3):983-1002

assignments tend to be fixed across all years of high school, resulting in long-term negative consequences of initial misassignments. They also charge that permanent assignments fail to take into account the differential rates of students' cognitive development.

Research on the effects of tracking on student achievement is fairly consistent in showing positive effects of tracking on the achievement of high-tracked students, negative effects on low-tracked students, and negligible effects on those assigned to the middle or regular track (see Slavin 1990, for a review of this literature). Other studies identify factors that affect initial track assignment; show that background characteristics other than ability influence track placement; and demonstrate that schools differ in the criteria they use for assignment (Hallinan 1991, 1992). In contrast to this large body of literature on track placement and its consequences, few studies examine the permanence of track assignments (exceptions are Rosenbaum 1976; Garet, Agnew & DeLany 1987; Pallas, Natriello & Riehl 1994). Most of the cross-sectional and longitudinal studies of tracking seem to assume that no changes in students' track level occur during or across school years.

It is important to know whether track assignments tend to be permanent in order to evaluate the effectiveness of tracking as a pedagogical practice. If students are not allowed to move from one track to another, then an inappropriate early track placement could have serious consequences for a student's achievement, especially since the effects would be cumulative over a student's high school career. If, on the other hand, students are reassigned to different tracks in response to changes in their achievement or when a misplacement is detected, then tracking may facilitate student learning.

Several questions arise concerning track mobility. The most fundamental one is whether track assignments are fixed across a school year or from one school year to the next. If students change track during their high school career, does the incidence of mobility vary by school and by subject area?

Questions also arise regarding the timing of track change. When during high school do track changes typically occur? Is change more likely at the beginning of high school when teachers do not know the students well and initial placement is based primarily on objective measures of ability? Or are track changes more frequent later in high school when students have fulfilled graduation requirements and are permitted to choose more elective courses? Are reassignments more likely to occur between school years than during an academic year?

Another set of questions concerns the direction of track mobility. When students change track, are they more likely to move to a higher or a lower track? How are students' ascribed and achieved characteristics related to the likelihood of their changing to a higher or lower track? Is the probability of an upward or downward move the same for mathematics as it is for English?

Further questions concern which students change track. Are track reassignments primarily a response to students' learning needs? Are nonacademic considerations taken into account in making track reassignments? Do students' background characteristics play a role in track changes? Are student reassignments influenced by structural and organizational characteristics of schools, such as school size, instructional resources, and teacher preferences?

Finally, one may ask whether track mobility accomplishes its presumed aim, namely, to increase track homogeneity. If track reassignments are governed primarily by the desire to improve the fit between student and level of instruction, then one would expect tracks to become more homogeneous over time. On the other hand, if other factors influence track change, such as student choice, and schedule conflicts, then track mobility may have little effect on track homogeneity or may result in greater track heterogeneity.

This article provides an in-depth analysis of track mobility for a sample of secondary school students. The research will document the frequency and direction of track change; examine the individual-level determinants of track mobility; and identify the factors that affect upward and downward track mobility as well as those that influence dropping a subject in which courses are tracked. The study also will demonstrate the effects of track mobility on track homogeneity.

Conceptualizations of Track Mobility

Track selection systems have been conceptualized in a number of ways. In an early essay, Turner (1960) describes two ideal types of social mobility patterns that are useful in characterizing track change: sponsored mobility and contest mobility. In sponsored mobility, individuals are selected at an early point in their lives for elite status. They are separated from those not selected, given special preparation, and provided with access to elite positions not offered to others. Turner argues that in British schools, streaming (tracking) is described by a sponsored mobility model. The sponsored mobility model implies virtually no track mobility, since those originally selected as elite generally maintain that status while those not selected cannot attain that status themselves.

Contest mobility, in contrast, denies no one access to elite positions. Indeed, under contest mobility, less successful individuals may be given remedial help to increase their chances of competing. Separation is delayed and its consequences minimized. Contest mobility never guarantees status security since newcomers can always displace those in status positions. Turner claims that in American schools, tracking resembles a contest mobility model. The contest mobility model permits movement in either direction, tied to meritocratic criteria, while encouraging upward mobility.

Rosenbaum (1976) proposes a different model of track mobility, namely a tournament model. In a tournament, winning grants the right to go on to the next round; losing results in elimination from the contest. In a case study of tracking in a junior and senior high school, Rosenbaum identifies three characteristics of the school's tracking system: little change into or out of the upper, college track, little movement into college tracks from noncollege tracks, and considerable movement into noncollege tracks from the lower college track. Based on these results, he claims that tracking is a tournament in that once a person moves out of a college track, the opportunity to return to that track level is lost permanently. Remaining in a college track provides better opportunities to learn and sustains the hope of future success. The tournament model, therefore, predicts only downward mobility, based on meritocratic standards.

Garet, Agnew and DeLany (1987) view track mobility as a matching problem. They argue that course selection decisions are made jointly by school personnel and students. Schools determine the structure of the track system, including which courses are offered and when they are scheduled; students choose courses from among these offerings. Schools are constrained by resources while students are influenced by college aspirations, peer networks, and scheduling constraints. Matching students to courses demands compromises by both the school and the student. The process resembles an optimization problem in which the number of school and student preferences satisfied is maximized. The optimization model presupposes mobility that is somewhat random with respect to students' background characteristics and academic ability.

Hallinan and Sorensen (1986) depict track mobility as a vacancy competition. Individuals can advance to a higher position in a structure only when one becomes vacant. When a position is available, it is offered to the candidate who ranks highest on the relevant criteria. In a tracking structure, a position in a track becomes available when a student in one track moves to a higher or lower track or leaves the track structure. The opportunity to move into the vacancy is offered to the student in the next lower track who ranks highest in the ability distribution of that track. School organizational characteristics, including teacher resources, space, and norms governing teaching loads determine the initial size of the tracks while various student and school factors affect the rate of vacancies and the amount of track movement. The vacancy competition model permits upward movement, based on meritocratic criteria.

Barr and Dreeben's (1983) conceptualization of the formation of ability groups in a first grade class suggests another mechanism governing track mobility, which can be referred to as a technical model. They claim that the number, size, and composition of ability groups in an elementary school classroom are a response to distributional characteristics of the class. If, for example, a large number of students are at the low end of the ability distribution, then teachers may form two low ability groups to accommodate their needs. If the ability distribution is fairly uniform, then equal-sized ability groups are likely to be created. In this framework, a teacher moves students to different ability groups in order to accommodate their learning needs. Applied to tracking in the high school, this perspective would require that schools take into account the ability distribution of the students in predetermining track size and that they, rather than the students, make decisions regarding track changes. The technical mobility model, like the contest mobility model, allows upward or downward mobility dependent on student performance.

The models predict varying degrees of change in track homogeneity over time. The contest and technical models imply increasing homogeneity in all tracks, since movement is dependent on student achievement. The tournament model predicts increasing homogeneity in the higher tracks and decreasing homogeneity in the lower tracks since only downward movement is permitted. The models of sponsored mobility, vacancy competition, and optimization imply little change in track homogeneity at any track level.

Characteristics of Track Mobility in High School

These various mobility models depict different kinds of track movement based on different motivations. The models are based on various assumptions about the rules and regulations governing the track selection process in contemporary American high schools. Some models assume that track decisions are made by school personnel without taking student choice into account. Others assume that students have a voice in the selection process. Some models present track change primarily as an effort to create more homogeneous instructional groups. Other models assume that school authorities take into account both student ability and other student characteristics. The fit between a model's assumptions and the rules and regulations that govern tracking in a high school determines the validity of the model as a representation of track mobility.

In most contemporary high schools, while school authorities aim to create homogeneous instructional groups, they employ a number of criteria in making track assignments. They consider standardized test scores as a measure of student ability and grades as a measure of student achievement. In addition, they consider teacher recommendations that convey teacher perceptions of student motivation, effort, and academic interests. Moreover, contemporary schools typically permit some degree of student choice in track selection, allowing students to base their choices on factors such as educational and occupational aspirations, peer influences, scheduling conflicts, and time constraints.

The process of track selection and mobility in the schools in this study is described in the student manual. The same rules govern track selection in many contemporary high schools. The manual states:

> Students are guided toward placement in specific courses based on criteria related to reading ability, grade record, previously completed courses, test results, and/or recommendations of teachers and counselors. Student or parent request for placement in particular course or level of a course will be fully considered.

Some of the proposed conceptual models of track mobility fail to depict the process of track change accurately because they fail to consider the various criteria schools use in making placement decisions and may ignore student choice as a factor in the process. For example, the sponsored mobility, tournament, and vacancy competition models place primary emphasis on student ability and do not consider student choice. In contrast, the optimization model emphasizes the interaction between school and student choices while the contest mobility model permits movement in any direction, based on unspecified criteria. Consequently, the sponsored mobility, tournament, and vacancy competition models, as well as Barr and Dreeben's distributional model, are not expected to describe track mobility in high school. The contest mobility and optimization models are more likely to depict patterns of track change.

In addition to the pattern of track change described by the contest mobility and optimization models, several other features of track mobility can be anticipated. The timing of track change is likely affected by two factors. First, track changes that occur during a school year disrupt the learning process. If students move to a higher track, they must make up material not covered in the

lower track. Regardless of whether a student moves to a higher or lower track, the teacher typically must devote some instructional time to help the student adjust, thereby depriving the other students of instructional time. In addition, the social relations and the routine of the classroom may be altered. Track change that occurs between school years is less disruptive and consequently more likely than change during the school year.

The timing of track change in a student's high school career depends on several factors. The initial 9th grade track assignment is based on objective measures of ability, recommendations, and parental and student preferences (Hallinan 1991). Since high school personnel generally do not know the incoming students personally, some misassignments are likely, necessitating track changes in 9th grade to correct them. Fewer misassignments would be anticipated in the upper grades, since counselors and other school personnel would know the students better and could make more accurate placements. At the same time, the curriculum is more complex in the latter years of high school, and students have more opportunities to select different courses in different tracks. The need to correct misassignments in the early high school grades and the number of course options available to students in the later grades lead one to expect some degree of track mobility throughout high school.

Graduation requirements affect the likelihood that a student leaves the tracking structure, a departure that can come about in two ways. First, if students feel unprepared for a particular tracked course, they may take an untracked remedial course to strengthen their abilities before enrolling in the more difficult tracked course later. Second, students may choose not to take any English or mathematics courses during a given year in order to take a preferred course. In this case, if the course omitted is English, the student must make up the tracked course either in summer school or by taking two English courses in a subsequent year, in order to earn the eight English credits required for a diploma. Since only four credits of mathematics are required for graduation, students may drop out of mathematics without jeopardy after two years.

A number of factors affect the direction of track mobility. Most of the conceptual models of mobility imply that movement to a higher track is less likely than change to a lower track. Nevertheless, the role of student choice in track mobility suggests a different pattern. Students may be attracted to higher track courses because of college aspirations, parental pressure, status, teacher reputation, or the curriculum. In addition, organizational characteristics of a track system, such as the elimination of low-track mathematics courses in eleventh or twelfth grade, increase the likelihood of some students' moving to a higher track. Students also are attracted to lower-track courses to avoid failure or to lessen study time. The various motivations that govern student track selection suggest that considerable upward and downward mobility can be expected.

Empirical research on gender and race biases in education alert one to the possibility that background characteristics may exert an influence on track mobility. To the extent that females in high school are more proficient in verbal skills but have less confidence in their mathematical abilities than males, as is suggested in the literature, they are more likely to change to higher-tracked English courses and lower-tracked mathematics courses than males. If teachers

associate the typically lower standardized test scores of black students with lower motivation and effort, they are likely to encourage blacks to move to lower-level courses more frequently than whites.

Since the aim of tracking is to create homogeneous ability groups, one would expect track change to lead to greater track homogeneity over time. However, other factors influencing track change make the attainment of this goal less likely. The reliance of teachers on a number of criteria in recommending track change and the role of student choice in tracking decisions suggest that increasing track homogeneity is not an operational goal in the schools. As a result of the various factors that affect track assignments, an increase in track homogeneity as a result of track mobility is not expected.

The research presented here documents the frequency of track change in English and mathematics, the direction and timing of track mobility, the characteristics of students who change track, and the consequences of track mobility for track homogeneity.

Methodology

SAMPLE

In 1986, a longitudinal survey of tracking in middle and secondary schools was initiated. Information was obtained for all the students beginning 7th grade that year in seventeen public and private elementary and middle schools in two midwestern cities. Similar information was gathered from the following cohort of 7th grade students the next year. In 9th grade, these two cohorts of students fed into six public high schools, two Catholic schools, and one private school. One of the Catholic high schools chose not to participate in the study, and since the private high school did not track students, it was also excluded from the analyses.

The research reported here is based on data from the two cohorts of students who attended one of the six public high schools or the one Catholic high school in the sample. Information was collected on the tracking structures and policies of the high schools in the sample. These data include the number and size of tracks, track composition with respect to gender, race, ethnicity, and academic ability, and criteria for assigning students to tracks. In addition, changes in student track assignments within and across academic years were recorded.

Students' gender, race, ethnicity, age, and free- or reduced-lunch status were obtained from student records. Because only low income students qualified for the free- or reduced-lunch program, these data are used as a measure of socioeconomic status. School records provided grades in English and mathematics at the end of each semester. Standardized scores in English and mathematics from the Indiana Statewide Test of Educational Progress (ISTEP), a mandatory statewide test administered to students in the 9th and 11th grades, also were obtained from school records.

VARIABLES

The dependent variable in the analysis is track mobility. Students may move from the track to which they are initially assigned or may retain their original placement. For each student in the sample, a complete record of initial track placements and every change in those placements throughout high school are available. Track mobility is coded as change = 1 and no change = 0.

The independent variables for the analyses include gender (female = 1, male = 0), race (black = 1, white [including a small number of Asians, Hispanics and American Indians] = 0), socioeconomic status (free lunch = 1, otherwise = 0), and age (in years). Ninth grade standardized test scores in English and mathematics, coded as English 9-ISTEP and math 9-ISTEP, provide a measure of ability. Achievement is measured by student grades in English and mathematics, coded as F = 0, D = 1, C = 2, B = 3, A = 4.

No school or track-level data are missing, since these data were obtained from district and school records and from principals, teachers, and counselors. A small amount of individual-level data are missing, due primarily to missing grades. This represents less than 2% of the sample. Students who were not part of the usual tracking structure during high school are excluded from the study, including special education, learning disabled, and bilingual students. Students who transferred out of or into the schools in the sample after the beginning of 9th grade or who dropped out of high school are excluded also. The transfers and dropouts represent about 10% of the student population. While the dropouts likely have lower test scores and grades than the students in the sample, this should not affect the track analyses reported here.

METHODS OF ANALYSIS

Both descriptive and inferential analyses are conducted. In the descriptive analysis, percentage distributions and records of track assignments illustrate the frequency, direction, and duration of track change over high school.

The inferential analysis relies on logistic regression models to examine the determinants of track mobility. Three models are estimated. The first model predicts the effects of the independent variables on the probability of moving to a higher track, compared to not changing track. The second model estimates the likelihood of changing to a lower track, compared to not changing track. The third model examines the determinants of dropping out of the track sequence in English and mathematics, compared to continuing to take tracked English or mathematics courses.

Results

Characteristics of the students in the total sample of 2,394 students in English and 2,391 students in mathematics are presented in Table 1. The sample is fairly evenly divided by gender. About 16% of the students are black and 16% have low socioeconomic status (SES). On the average, the students are slightly below 15 years old in the middle of 9th grade. The average grade in English and mathematics is C+. Approximately 6% of the sample is in each of the basic and

TABLE 1: Means and Standard Deviations of Variables for Pooled Sample of Two Cohorts[a]

	Mean	S.D.	N	
Female	.51	.50	2,394	Dummy variable: Female = 1
Black	.16	.37	2,394	Dummy variable: Black = 1
Age	-.12	.45	2,394	Age (measured in years) minus 15
Free lunch	.16	.36	2,394	Dummy variable: ever receive free lunch = 1
Cohort 2	.49	.50	2,394	Dummy variable: Cohort 2 = 1
English 9-ISTEP	62.88	22.76	2,393	ISTEP, 9th grade mean English percentile
Math 9-ISTEP	65.35	24.11	2,390	ISTEP, 9th grade mathematics percentile
English 9 grade	2.53	.98	2,382	Final 9th grade English grade (A = 4)
Math 9 grade	2.28	1.14	2,369	Final 9th grade Mathematics grade (A = 4)
Basic English 9	.06	.23	2,394	Dummy: 9th grade English track, Basic = 1
Regular English 9	.51	.50	2,394	Dummy: 9th grade English track, Regular = 1
Honors English 9	.37	.48	2,394	Dummy: 9th grade English track, Honors = 1
Advanced English 9	.06	.24	2,394	Dummy: 9th grade English track, Advanced = 1
Very basic math 9	.10	.30	2,391	Dummy: 9th grade math track, very basic = 1
Basic math 9	.17	.38	2,391	Dummy: 9th grade math track, Basic = 1
Regular math 9	.41	.49	2,391	Dummy: 9th grade math track, Regular = 1
Honors math 9	.23	.42	2,391	Dummy: 9th grade math track, Honors = 1
Advanced math 9	.09	.29	2,391	Dummy: 9th grade math track, Advanced = 1
School 1.2	.13	.33	2,394	Dummy: School 1.2 = 1
School 1.3	.17	.37	2,394	Dummy: School 1.3 = 1
School 1.4	.13	.34	2,394	Dummy: School 1.4 = 1
School 1.5	.13	.33	2,394	Dummy: School 1.5 = 1
School 2.1	.17	.38	2,394	Dummy: School 2.1 = 1
School 3.1	.14	.35	2,394	Dummy: School 3.1 = 1
Upward mobility in English	.19	.39	2,394	Dummy: English track upwardly mobile = 1
Downward mobility in English	.11	.31	2,394	Dummy: English track downwardly mobile = 1
Drop track English	.07	.26	2,394	Dummy: No English tracked class = 1
Upward mobility in math	.07	.26	2,391	Dummy: Math track upwardly mobile = 1
Downward mobility in math	.04	.20	2,391	Dummy: Math track downwardly mobile = 1
Drop track math	.64	.48	2,391	Dummy: No English tracked class = 1

[a] The sample contains those students in the same high school or school district for grades nine through twelve and initially assigned to a 9th grade English or mathematics track.

advanced English tracks, 51% is in the regular track, and 37% is in the honors track at the beginning of 9th grade. In mathematics, 10% of the students are in the very basic track, 17% in the basic, 41% in the regular, 23% in the honors, and 9% in the advanced track.

The analyses address the following five questions: (1) How much track mobility occurs during high school? (2) When does track mobility occur? (3) Do students tend to move to higher or lower tracks, or to discontinue taking a tracked subject? (4) Who is most likely to move? (5) Does mobility increase track homogeneity?

HOW MUCH TRACK MOBILITY OCCURS?

The descriptive statistics in Table 1 show that 30% of the students (upward mobility in English + downward mobility in English) change English track during high school and 11% of the students (upward mobility in math + downward mobility in math) change mathematics track. Nineteen percent of the students complete high school at a higher English track than the one to which they were originally assigned in 9th grade, while 11% finish in a lower track, and 7% drop tracked English. In mathematics, 7% finish high school at a higher track and 4% at a lower track than they started, while 64% drop tracked mathematics.

The data reveal a considerable amount of track mobility during high school. Almost one-third of the students finish high school in a different English track than the one in which they started. Less mobility is found in mathematics, partly because stringent prerequisites for movement to upper-level courses prevent some students from moving up. Overall, the data refute the claim that tracking locks students into a track level at the beginning of high school and precludes subsequent change.

Only a small number of students drop out of tracked English courses during high school because four years of English are required for graduation. Those who drop an English course one year need to make it up in summer school or by taking two English classes in a subsequent year. A large percentage of students drop mathematics before the end of twelfth grade, likely because only two years of mathematics are required.

WHEN DO STUDENTS MOVE?

To better understand track mobility, it is necessary to determine when track changes occur. Table 2 shows that far more track mobility occurs between school years than during the year. For example, 14% of the students change English track between the beginning of 9th grade and the beginning of 10th grade, and most of this change takes place over the summer. The same pattern is found in subsequent years. Similarly, more students change mathematics track over the summer than during the school year. In other words, students tend to move to a higher or lower track after a specific course has been completed rather than while it is in progress, thereby minimizing disruption of instruction.

TABLE 2: Proportion of Students Who Change Track by Year in School[a]

	English		Mathematics	
	p	N	p	N
Begin grade 9 to end 12	.40	2,160	.47	825
Begin grade 9 to end grade 9	.05	2,391	.12	2,381
End grade 9 to begin grade 10	.11	2,382	.13	2,158
Begin grade 9 to begin grade 10	.14[b]	2,380	.23	2,156
Begin grade 10 to end grade 10	.02	2,376	.04	2,121
End grade 10 to begin grade 11	.11	2,348	.18	1,725
Begin grade 10 to begin grade 11	.13	2,343	.21	1,721
Begin grade 11 to end grade 11	.04	2,334	.07	1,622
End grade 11 to begin grade 12	.19	2,265	.18	1,042
Begin grade 11 to begin grade 12	.21	2,243	.24	1,029
Begin grade 12 to end grade 12	.04	2,202	.02	845

[a] The sample contains those students in Table 1 who were assigned to an English or Mathematics track at the beginning and end of any included school year.
[b] Some students changed tracks both between the beginning of 9th grade and the end of 9th grade and between the end of 9th and the beginning of 10th. Hence, the annual statistics show fewer moves than the sum of the school year and summer counts. The same is true for any short-term/long-term comparison of proportions.

At the same time, the amount of mobility that occurs during the school year is not negligible, particularly in mathematics. During 9th grade, 12% of the students in the seven schools in the sample, or 286 students, change to a different mathematics track. Depending on whether these changes result in better placement, they may or may not be educationally beneficial either to the student who moves or to the student's peers.

A second pattern in the timing of track mobility is that more change occurs later in students' high school careers than earlier. Table 2 shows that in English, 21% of the students change track between 11th and 12th grade, compared to only 14% between 9th and 10th, and 13% between 10th and 11th. In mathematics, about 24% of the students change track between 11th and 12th grades, compared to considerably smaller percentages between 9th and 10th grades (23%) and between 10th and 11th grades (21%).

The pattern of greater track change in the upper grades is due to several factors. First, students can no longer take very basic mathematics after 10th

grade, or basic mathematics after 11th grade. Second, while students in lower grades take mostly required courses, those in upper grades take a number of electives that may conflict with tracked mathematics and English courses, leading them to change those courses. Finally, since older students are more likely to work after school, those in upper tracks may select easier courses to avoid having to spend long periods of time on homework.

WHERE DO STUDENTS MOVE?

Critical to understanding the nature of tracking is a determination of where students go when they change track. Table 1 shows that students in the English track moved mostly within the track structure and students in the mathematics moved mostly out of the track structure. The data (not presented) show that about two-thirds of the students who remain in tracked English and mathematics classes finish high school in a higher track than they started in, while one-third complete high school in a lower track.

One other move is possible within the track structure. Students may change to a higher or lower track during high school and later return to their original track. Since this is a somewhat small sample of students, it permits an examination of the direction and duration of these track changes. The data (not presented) reveal that most of the students remain out of their originally assigned track for only one or two semesters. For example, 18 students originally assigned to the regular English track move down to the basic track for one semester and nine more move down to the basic track for two semesters before returning to the regular track. The remaining six students who move down from the regular track continue in the basic track for a longer period of time. Similarly, of the 123 students who move up one or more tracks in English, most move from the regular track to the honors track for only one or two semesters.

A similar tendency toward short-term change is found in mathematics. Fourteen of the 39 students who move from the honors to the regular track return to the honors track after one or two semesters. Fourteen other students who make this change remain in the honors track for five semesters. However, 13 of these students are from school 2.1, suggesting the occurrence of an idiosyncratic school change. The most common upward change in mathematics is from the regular track to the honors track. Of the 129 students who change to a higher mathematics track, 66 students move from the regular to the honors track for one or two semesters and then return to the regular track. The short-term nature of most of these track changes suggests that students are exploring different track levels but finding that their initial placement was appropriate.

WHO MOVES?

To identify the predictors of track mobility, two logistic models are estimated: a model comparing change to a higher track to no change and a model comparing change to a lower track to no change.

Table 3.1 presents the estimates of a logistic model that predicts the likelihood of moving to a higher track as opposed to not changing track, on the basis of student characteristics, track level, and school. The results show that

TABLE 3: Logistic Regressions Predicting Upward Track Mobility[a]

	English		Mathematics	
Variable	B	S.E.	B	S.E.
Female	-.19	.13	.54*	.22
Black	-.09	.20	-.21	.46
Age	.35*	.16	.09	.32
Free lunch	.40+	.21	.02	.52
Cohort 2	.14	.12	-.20	.22
Grade 9 ISTEP	-.04**	.00	-.04**	.01
Grade 9	-.37**	.08	-.68**	.15
Basic (Grade 9)	-3.31**	.29		
Regular (Grade 9)	Comparison		-2.10**	.28
Honors (Grade 9)	.86**	.16	Comparison	
School 1.2	-.14	.20	1.43**	.43
School 1.3	.39*	.20	1.37**	.49
School 1.4	.44+	.23	.03	.45
School 1.5	.36+	.21	-.43	.49
School 2.1	.78**	.22	-.27	.41
School 3.1	3.74**	.41	-.31	.36
Constant	3.90**	.33	7.26**	.94
-2 Log likelihood	1,656.33		546.02	
N	1,836		597	

[a] A negative coefficient indicates a greater likelihood of changing track versus not changing track.

+ $p < .10$ * $p < .05$ ** $p < .01$

older students are less likely to move to a higher English track than younger students. Students who have higher standardized test scores and grades in English are more likely to move to a higher track.

In addition, students initially assigned to the basic track are more likely to move to a higher track compared to those in the regular track, while students in the honors track are less likely to move up than those in the regular track. Those in the advanced track are excluded from this analysis because they cannot move to a higher track. Students in schools 1.3, 2.1, and 3.1 are signifi-

cantly less likely to move to a higher track than those in school 1.1, while schools 1.4 and 1.5 show a similar though weaker tendency. The greater tendency of students in school 1.1 to be upwardly mobile in English is interesting because this school has the strongest academic reputation among the public schools in the sample. It is likely that a strong academic climate motivates upward track mobility.

In mathematics, the model shows that females are less likely to move to a higher mathematics track than males. Higher ISTEP scores and course grades are associated with greater upward change. Students in the regular track are more likely to move up than those in the honors track. Basic track students are excluded from the analysis because they must move up or out for structural reasons since the basic track is eliminated after 11th grade. Those in the advanced track are excluded because they cannot move to a higher track. Students who attend schools 1.2 and 1.3 are less likely to move to a higher track than those in school 1.1.

These results show that strong performance leads to upward track mobility. However, this effect is constrained by track placement. Controlling for ability and achievement, we find that students assigned to a lower track are more likely to move to a higher track than their peers in higher tracks.

Table 3.2 shows the predictors of moving down one or more tracks compared to not changing tracks. Students are more likely to move to a lower English track if they are male, have low income, low standardized test scores and grades, or are in the honors or advanced tracks, compared to the regular track. Students in schools 1.2 and 3.1 are significantly more likely to move to a lower track than those in school 1.1. In mathematics, low standardized test scores and low grades lead to a downward track change. Students who attend school 2.1 are significantly less likely to move to a lower track in mathematics than those in school 1.1.

In general, weak performance predicts downward track mobility. However, as with upward mobility, track level also has a direct effect on downward mobility. Assignment to a higher track increases the likelihood of moving to a lower track, regardless of ability and achievement.

These analyses reveal that student background, ability, and achievement affect track mobility. Moreover, track assignment exerts an independent effect on the likelihood of a track change and the direction of the change. As a strong predictor of track mobility, track placement plays a crucial role in providing students with opportunities to learn.

Since a considerable number of students discontinue taking both English and mathematics during high school, a third logistic model examines the determinants of dropping out of a tracked English or mathematics course sequence. This analysis permits a comparison of the factors that affect upward and downward track mobility in English and mathematics with the predictors of discontinuation in these subjects.

Table 4 presents the determinants of dropping out of the English or mathematics track sequence. The parameter estimates show that females, those with lower grades, and, to a lesser extent, low-SES students are more likely to drop English. Moreover, students in the basic track are more likely to drop English compared to those in the regular track, while those in the higher tracks

TABLE 4: Logistic Regressions Predicting Downward Track Mobility[a]

	English		Mathematics	
Variable	B	S.E.	B	S.E.
Female	.36*	.17	-.21	.27
Black	.00	.29	.08	.57
Age	-.32	.22	-.33	.40
Free lunch	-.84**	.28	.06	.61
Cohort 2	.19	.17	.37	.27
Grade 9 ISTEP	.04**	.01	.07**	.01
Grade 9	.88**	.12	.92**	.17
Regular (Grade 9)	Comparison			
Honors (Grade 9)	-5.26**	.34	Comparison	
Advanced (Grade 9)	-6.36**	.50	-.58+	.32
School 1.2	-.92**	.35	-.05	.44
School 1.3	-.07	.35	-.70	.46
School 1.4	.00	.35	.73	.53
School 1.5	-.70+	.38	.26	.53
School 2.1	.67+	.36	3.46**	1.08
School 3.1	-3.35**	.37	.62	.48
Constant	1.32**	.47	-7.80**	1.31
-2 Log likelihood	931.56		369.87	
N	1,713		472	

[a] A negative coefficient indicates a greater likelihood of changing track versus not changing track.

+ $p < .10$ * $p < .05$ ** $p < .01$

are less likely to drop it. A school effect is observed also; students in schools 1.2, 1.4, 1.5, and 2.1 show a stronger tendency to drop English than those in school 1.1.

Females, older students, and low-SES students are more likely to drop mathematics than their peers. Whites are slightly more likely to drop mathematics than blacks, as are students with low test scores and grades. Pupils in schools 1.2 and 3.1 are less likely to drop mathematics than those in school 1.1. These findings show that poor performance increases the likelihood of dropping

TABLE 5: Logistic Regressions Predicting Outward Track Mobility[a]

	English		Mathematics	
Variable	B	S.E.	B	S.E.
Female	-.56**	.19	-.37**	.13
Black	-.29	.27	.45+	.25
Age	-.16	.19	-.56**	.19
Free lunch	-.40+	.23	-.84**	.25
Cohort 2	.43*	.18	.29*	.13
Grade 9 ISTEP	-.00	.01	.01**	.01
Grade 9	.59**	.11	.58**	.08
Basic (Grade 9)	-1.07**	.36		
Regular (Grade 9)	Comparison		-1.10**	.17
Honors (Grade 9)	.96**	.28	Comparison	
Advanced (Grade 9)	1.98+	1.06	.60**	.22
School 1.2	-1.27*	.60	1.99**	.25
School 1.3	-.54	.60	-.11	.24
School 1.4	-1.63**	.56	-.03	.26
School 1.5	-1.49**	.56	-.51+	.27
School 2.1	-3.22**	.54	-.32	.24
School 3.1	.46	.79	1.38**	.24
Constant	2.71**	.64	-2.87**	.47
-2 Log likelihood	834.33		1,426.05	
N	1,676		1,486	

[a] A negative coefficient indicates a greater likelihood of changing track versus not changing track.

+ $p < .10$ * $p < .05$ ** $p < .01$

out of the mathematics track structure. Assignment to a lower track also increases the likelihood of dropping out of tracked mathematics courses, regardless of ability and achievement.

Comparing the determinants of track mobility in English and mathematics with the factors that affect dropping these subjects shows that lower grades and test scores are the strongest predictors of both moving to a lower track and dropping a tracked subject. In examining the influence of ascribed characteristics on course choices, the data show that females are less likely than males to move

to a higher mathematics track, and, although they are less likely to move to a lower track, they are significantly more likely to drop mathematics. Race has little effect on track mobility but whites are slightly more likely than blacks to drop mathematics. Older students are less likely to move to a higher English track and more likely to drop mathematics. Low-income students are slightly less likely to change to a higher-English track, more likely to change to a lower-English track, and more likely to drop English and mathematics than older students.

In general, the comparisons show that ascribed characteristics affect dropping out of a tracked course sequence in the same way they influence track mobility. Hence, being female, older, and having a low income appear to place students at risk for changing to lower tracks and dropping out of tracked subjects. In this way, tracking can be seen as a mechanism through which fewer opportunities for learning are channeled to these students.

DOES TRACK MOBILITY INCREASE TRACK HOMOGENEITY?

If track changes are governed primarily by an effort to improve the fit between a student's learning ability and the level of instruction provided at a particular track level, then mobility should result in increased track homogeneity. Table 5 presents the standard deviations of the ISTEP scores of the students at the beginning and end of each year in high school by track and subject for school 1.1. This school was chosen for illustration because it serves as the comparison case in the logistic analyses. The results are similar for all the other schools in the sample, including the private school.

The data in Table 5 show little change in track homogeneity in English or mathematics over the school year at any track level. Some exceptions are when the size of a small track changes, as in basic English 11, or when there is a considerable change in the size of a large track, as in advanced math 9. In these two cases, track homogeneity increases markedly. A more impressive example is basic math 9 where track homogeneity increases after fourteen new students are added to the track. In general, student mobility has little effect on the distribution of achievement in a track. Clearly, mobility is not being used in a systematic way by the schools to increase track homogeneity. This is likely due, to a large degree, to permitting student choice to play a significant role in track selection and change.

Conclusions

The pattern of track change observed in this study suggests that track mobility in American high schools has some of the characteristics of a contest mobility model. Students moved to both higher and lower tracks, but moved to higher tracks with somewhat greater frequency. Only the most advanced tracks had strict prerequisites that limited admission. The upward bias is evidence that organizational constraints on mobility play only a small role in track change and that lower-track students may be given special assistance in moving to higher tracks. The fact that students participate in the track selection process requires a modification of the contest mobility model as a description of track mobility.

TABLE 6: Standard Deviations of ISTEP Scores at Beginning and End of Each School Year by Track for School 1.1

	English							
	Basic		Regular		Honors		Advanced	
Grade/Time	S.D.	N	S.D.	N	S.D.	N	S.D.	N
English ISTEP 9 Begin	9.44	10	18.22	184	15.15	98	5.85	35
English ISTEP 9 End	9.32	13	18.35	188	14.17	87	6.13	37
English ISTEP 10 Begin	10.14	13	18.09	184	13.74	97	5.61	31
English ISTEP 10 End	10.46	10	18.47	192	13.58	94	5.61	31
English ISTEP 11 Begin	10.51	6	18.99	162	15.25	120	5.93	33
English ISTEP 11 End	3.09	4	18.76	169	14.34	114	5.93	33
English ISTEP 12 Begin	13.33	3	18.63	135	17.65	116	14.39	68
English ISTEP 12 End	9.60	5	18.45	128	18.58	122	14.72	65

	Mathematics									
	Very Basic		Basic		Regular		Honors		Advanced	
Grade/Time	S.D.	N	S.D.	N	S.D.	N	S.D.	N	S.D.	N
Math ISTEP 9 Begin	15.27	29	21.54	33	16.31	158	11.67	71	8.43	33
Math ISTEP 9 End	17.13	30	19.44	47	15.41	156	8.08	66	2.76	23
Math ISTEP 10 Begin	13.15	15	20.44	55	16.00	144	10.98	65	3.01	25
Math ISTEP 10 End	12.58	17	20.12	55	15.86	141	10.89	61	2.99	26
Math ISTEP 11 Begin	No very		19.89	34	16.26	117	15.74	86	2.76	24
Math ISTEP 11 End	basic track in 11th		19.87	35	15.63	108	13.92	73	2.73	24
Math ISTEP 12 Begin	or 12th		No basic		17.24	68	11.53	35	10.75	34
Math ISTEP 12 End			track in 12th		16.09	37	10.11	29	12.34	39

Since schools must consider the preferences of the entire student body and match them to course offerings, compromises must be made by both school personnel and students. This matching superimposes an optimization process on the mobility patterns. Track mobility in the schools examined in this study can best be described as a process that allows fairly unrestricted movement across tracks, constrained by the general goal of maintaining somewhat homogeneous instructional groups and by the desire to satisfy the preferences of as many teachers and students as possible. That is, track change resembles a contest mobility model constrained by efforts to optimize choices.

A major conclusion of this study is that track mobility occurs more frequently in high school than is generally believed. Students changed tracks in English more than in mathematics, likely because of stricter course prerequisites limiting upward mobility in mathematics. Most of the track changes in both English and mathematics occurred in the later grades, although a substantial number of students moved during 9th and 10th grades as well.

Some of the observed track changes undoubtedly can be attributed to efforts by school personnel to create a better fit between student ability and course content. School authorities make track decisions based on measures of student ability and judgments about student motivation. Several additional bases for track mobility are introduced when students participate in the decision process, some of which may be unrelated to academic considerations. The data presented here show that track changes do not necessarily result in more homogeneous tracks over time. To increase the effectiveness of track mobility in creating more homogeneous tracks, nonacademic motivations for track change need to be minimized. Further, it may be necessary for counselors to retain primary decision making power in approving track changes.

The study reveals gender, race, and income effects on track changes. Patterns of track mobility show that females are less likely to move to a higher mathematics track than males, while males have likely to change to a lower English track than females. Black students are more likely to drop mathematics than whites. Low-income students are more likely to move to a lower English track or to drop English or mathematics than higher-income students. These findings point to the need for close scrutiny of decisions governing track change to insure that track decisions channel equal opportunities for learning to students who differ in background characteristics.

In general, this study provides convincing evidence that track assignments are not fixed, as commonly believed. A surprising amount of track change occurs during each year of high school. At the same time, the patterns of track mobility observed in this study reveal a different source of inequity associated with tracking, namely an effect of background on the direction of track change. Schools must exercise considerable care in monitoring tracking assignments to avoid a relationship between student social characteristics and track decisions. Track mobility is a complex process that can have powerful, though often unrecognized, effects on student learning.

References

Barr, Rebecca, and Robert Dreeben. 1983. *How Schools Work*. University of Chicago Press.

Garet, Michael, John Agnew, and Brian DeLany. 1987. "Moving Through the System: Curriculum Decision Making in High Schools." Unpublished manuscript.

Hallinan, Maureen T. 1991. "School Differences in Tracking Structures and Track Assignments." *Journal of Research on Adolescence* 1:251-75.

____. 1992. "The Organization of Students for Instruction in the Middle School." *Sociology of Education* 65:114-27.

Hallinan, Maureen T., and Aage B. Sorensen. 1986. "Student Characteristics and Assignment to Ability Groups: Two Conceptual Formulations." *Sociological Quarterly* 27:1-13.

Pallas, Aaron, Gary Natriello, and Carolyn Riehl. 1994. "Tweaking the Sorting Machine: The Dynamics of Students' Schedule Changes in High School." Paper presented at the annual meeting of the American Sociological Association, Los Angeles.

Rosenbaum, James. 1976. *Making Inequality: The Hidden Curriculum of High School Tracking*. Wiley.

Slavin, Robert. 1990. "Ability Grouping in Secondary Schools: A Best-evidence Synthesis." *Review of Educational Research* 60:471-99.

Turner, Ralph. 1960. "Modes of Social Ascent through Education: Sponsored and Contest Mobility." *American Sociological Review* 25:855-67.

Conservative Protestantism and the Parental Use of Corporal Punishment*

CHRISTOPHER G. ELLISON, *University of Texas at Austin*
JOHN P. BARTKOWSKI, *University of Texas at Austin*
MICHELLE L. SEGAL, *University of Texas at Austin*

Abstract

The present study develops arguments linking Conservative Protestant affiliation and conservative beliefs about the Bible with the frequency with which physical punishment is used to discipline toddlers and preschoolers (ages 1-4) and older children (ages 5-11) and explores these ideas using data from the 1987-88 National Survey of Families and Households (NSFH). Multivariate results generally confirm that parents with conservative scriptural beliefs use corporal punishment more frequently than parents with less conservative theological views. Some modest net effects of Conservative Protestant affiliation are also observed. The study identifies several promising directions for future research on religious variations in child discipline.

The use of corporal punishment to discipline children has been the focus of perennial popular and academic controversy (see Larzelere 1994; Straus 1994a, 1994b). Although physical punishment is strongly supported and widely used by American parents (Straus & Gelles 1986; Wauchope & Straus 1990), critics claim that corporal punishment trains children in violence (1) by teaching that it is appropriate for powerful persons to subordinate others by using physical force, (2) by teaching that feelings of frustration and anger justify the use of violence, and (3) by establishing an association in a youngster's mind between love and violence (Gray 1988; Maurer 1974; Straus 1991). Consistent with such arguments, some research indicates that children subjected to corporal punishment tend to be more aggressive with peers and others than children who experience alternative forms of discipline (Hotaling, Straus & Lincoln 1990; Straus 1991). There is also evidence that persons who are spanked or slapped frequently in childhood may be more likely to use corporal punishment and to

** Portions of this study were presented at the 1994 meetings of the Association for the Sociology of Religion, Los Angeles, and the Society for the Scientific Study of Religion, Albuquerque. The authors thank Norval Glenn, George Holden, Paula Nesbitt, Daniel Powers, and two anonymous reviewers for helpful comments. The authors are solely responsible for the analyses and interpretations presented here. Direct correspondence to Christopher G. Ellison, Department of Sociology, 336 Burdine Hall, University of Texas at Austin, Austin, TX 78712-1088.*

 Social Forces, March 1996, 74(3):1003-1028

engage in harsh parenting and domestic abuse as adults (Egeland, Jacobvitz & Papatola 1987; Simons et al. 1991; Simons et al. 1993). In addition, even mild to moderate corporal punishment can sometimes cause serious physical injuries to young children (Taylor & Maurer 1985), and incidents that begin as mild punishment can escalate into abusive encounters. Further, some researchers have related harsh forms of physical punishment to negative developmental outcomes, including delinquency, psychopathology, academic failure, substance abuse, and others (Eckenrode, Laird & Doris 1993; Maccoby & Martin 1983; Wolfe 1987). Indeed, some studies suggest that the experience of corporal punishment even inclines children toward criminal violence in adulthood (Straus 1991; Widom 1989).

Given these and other concerns, researchers have worked to identify the social antecedents of corporal punishment. For instance, many have hypothesized (1) that working-class parents support and use corporal punishment more than their middle- and upper-class counterparts (e.g., Bronfenbrenner 1958), and (2) that African American parents support and use corporal punishment more than their white counterparts (e.g., Alvy 1987). To date, however, the empirical evidence on both points remains decidedly mixed (Cazenave & Straus 1990; Duvall & Booth 1979; Erlanger 1974).

Until recently, one important source of cultural support for corporal punishment received short shrift from researchers: *religion*. In an early exception to this general pattern of neglect, Erlanger (1974) analyzed retrospective data collected in the 1960s, and he noted that Baptists (1) were more likely to report that they had been spanked often as children, and (2) as adults, were more likely to approve of spanking in general than other respondents. Although these religious differences were substantial in magnitude, they received only brief discussion in that study — which was devoted mainly to social class differences in corporal punishment — and they were largely ignored by subsequent researchers.

Since the publication of Erlanger's (1974) article, several studies have renewed scholarly focus on religious variations in attitudes toward corporal punishment, with particular attention to the distinctive views of Conservative Protestants.[1] For instance, analyzing data on a nonrandom sample of church members from portions of Kentucky and Ohio, Wiehe (1990) reported that individuals affiliated with Conservative Protestant churches are more likely to view corporal punishment as an appropriate response to child misbehavior than are members of more liberal religious groups. Although Wiehe did not gauge the theological beliefs of his respondents directly, he argued that the churches that he classified as conservative also tend to endorse literalist or inerrantist views of the Bible, and he suggested that these beliefs contribute to pro-corporal punishment attitudes.

Following a similar line of argument, Grasmick and colleagues (Grasmick, Bursik & Kimpel 1991; Grasmick, Morgan & Kennedy 1992) used data on samples of Oklahoma City adults to show that members of Conservative Protestant denominations are especially inclined to favor corporal punishment in homes and schools. In one of these studies, Grasmick, Bursik, and Kimpel (1991) examined the scriptural beliefs of their respondents directly, and they concluded that high levels of Conservative Protestant support for corporal

punishment are largely accounted for by the tendency of Conservative Protestants to embrace literalist interpretations of the Bible.

Analyzing data from the 1988 General Social Surveys, Ellison and Sherkat (1993a) also reported that the members of Conservative Protestant denominations supported corporal punishment more strongly than other Americans. Moreover, they found that the relationship between Conservative Protestant denominational affiliation and support for corporal punishment is mediated by three key theological tenets: (1) the belief that the Bible should be interpreted as the literal Word of God; (2) the belief that human nature is fundamentally sinful and corrupt; and (3) the belief that persons who violate God's rules must be punished. In another study, Ellison and Sherkat (1993b) showed that this complex of theological beliefs also accounts for the disproportionate Conservative Protestant support for "authority-minded" parental values — i.e., the tendency to emphasize obedience in child rearing.

Although these recent investigations have shed new light on a neglected topic, they focus exclusively on *attitudes* toward corporal punishment among the general public. Consequently, they provide no direct evidence that Conservative Protestant parents themselves actually *use* physical discipline more often than other parents. Our study augments the literature in this area in several ways. We begin by reviewing "insider documents" on the topic — popular child-rearing manuals by conservative religious commentators — to develop a series of theoretical arguments linking Conservative Protestantism with corporal punishment. We then explore these ideas using data from the 1987-88 National Survey of Families and Households (hereafter NSFH). Our work centers on the parents of toddlers and preschoolers (children aged 1-4), and the parents of grade school-aged children (children aged 5-11). We consider the influence of several sets of factors — Conservative Protestant affiliation and conservative theology, parental and household characteristics, child characteristics and behaviors, and parental child-rearing values — on the frequency with which parents use corporal punishment to discipline specific children. In the concluding section of the article, we discuss several promising directions for future research on the links between Conservative Protestantism and child discipline.

Conservative Protestant Theology and Child Discipline

Why might Conservative Protestant parents use corporal punishment more frequently than other parents? Following the methodological counsel of McNamara (1985), we turn to an understudied set of insider documents, the substantial body of popular child-rearing manuals produced and distributed primarily within Conservative Protestant communities since the late 1960s (for details on this literature, see Bartkowski & Ellison 1995). Best-selling contributors to this literature include such conservative luminaries as James Dobson (1970, 1976, 1987), founder of the Focus on the Family organization, Beverly LaHaye (1977), founder of Concerned Women for America, noted conservative Christian radio television evangelist Chuck Swindoll (1991), and Marlin Maddoux (1992), president of the USA Radio Network, among many others.[2]

A careful review of these popular Conservative Protestant manuals reveals a clear and generally coherent theological rationale underlying support for the

corporal punishment of children (see Ellison & Bartkowski 1996; Ellison & Sherkat 1993a). Conservative religious beliefs about the right and responsibility of parents to discipline children with physical force commonly begin with the view that the Bible is the inerrant Word of God, the ultimate source of authority and guidance regarding every aspect of human life, including child rearing (Dobson 1970:197 and 1976:234-35; Fugate 1980:262-63; LaHaye 1977:145). Religious scripture is widely regarded as purposive divine communication with humanity, and therefore it is often presumed to contain reliable and sufficient information to guide the conduct of all human affairs. However, the implications of the doctrine of inerrancy for families, politics, and other human endeavors are perennially contested among Conservative Protestants (e.g., Barnhart 1993; Lindsell 1976). The theological and social meanings associated with biblical inerrancy are debated, produced, and disseminated by way of interpretive communities, or loose networks of theologians, pastors, and elite laity who share fundamental assumptions about the text (see Boone 1989).

The doctrine of biblical inerrancy is crucial for several reasons. First, contemporary constructions of inerrancy typically emphasize the ubiquity and divinely ordained nature of spiritual and worldly authority relations and articulate the principle that Christians should submit themselves to the leadership of duly constituted authorities in every area of life — church, family, school, workplace, and government (Ammerman 1987; McNamara 1985; Rose 1988). However, perhaps nowhere is this "authority-mindedness" (Wald, Owen & Hill 1989) clearer than in the popular literature on child rearing and family life. According to Conservative Protestant authors, families are characterized by specific patterns of authority relations (superordinate and subordinate roles) that should always remain unchallenged (Bartkowski & Ellison 1995).

The primary responsibility of children is to obey parental directives; in turn, parents are admonished to fulfill their obligations as authority figures rather than abrogating their leadership role within the family (see Christenson 1970; Fugate 1980). In support of these views, Conservative Protestant commentators marshal an impressive array of scriptural passages in which children are expected to honor and obey parental authority (Exodus 20:12; Ephesians 6:1-2; 1 Timothy 3:4-5), under threat of divine judgment (Exodus 21:15-17; Deuteronomy 21:18-21; Proverbs 30:17).

In addition to this preoccupation with themes of hierarchy and authority relations, contemporary constructions of biblical inerrancy draw heavily on Old Testament passages to emphasize themes of human sin and punishment (see Ellison & Sherkat 1993a, 1993b). Conservative Protestant commentators on child rearing argue that all individuals are born sinful (Psalms 51:5, 58:3) — that is, predisposed toward selfish conduct and inclined to reject authority in all forms (Christenson 1970:95-98; LaHaye 1977:3; Maddoux 1992:19-20).[3] For instance, Chuck Swindoll (1991) encourages parents to combat the "foolishness . . . bound up in the heart of a child" (Proverbs 22:15) with "diligent discipline" (Proverbs 13:24):

> "Foolishness" sounds rather mischievous and impish . . . somewhat lighthearted and fun-loving. But the Hebrews saw it as far more serious. "The fool has said in his heart 'There is no God'" (Psalm 14:1). The "foolish" possess a God-mocking, instruction-hating nature. Fools, therefore, despise discipline. Foolishness has a disrespect for authority.

Determined to go its own way, it resists all reproof. And remember, all this "is bound up in the heart of a child" — *your* child (88, emphasis in original).

Beverly LaHaye (1977) sounds an even shriller alarm:

[God] gave each of us a free will to choose evil or good, and the child that is not trained to choose good will undoubtedly choose evil. . . . It is of great benefit to the parent when he realizes that it is natural for his child to have a desire for evil. The child is not just being obstinate and uncooperative but is following that natural desire to learn more about and to experience *evil* (3, emphasis in original).

According to Conservative Protestant parenting specialists, "willful defiance" is virtually inevitable among children, particularly toddlers, and such defiance carries dangerous long-term consequences if unchecked. They believe that without appropriate discipline, children will grow up lacking respect for parents and other authority figures, nurturing rebellious tendencies that will make them unfit citizens, unproductive workers, and unhappy spouses and parents (Christenson 1970:88-90; Dobson 1970:14; Fugate 1980:14-15; LaHaye 1977:2).

Although Conservative Protestant child-rearing specialists extol the value of stern discipline — and particularly corporal punishment — on instrumental or pragmatic grounds, for securing behavioral conformity and training children for adulthood, most also believe that this form of punishment conveys an important religious lesson to youngsters about the nature of God. In brief, they argue that parents symbolize God's authority to children, and that many children come to conceptualize or envision God, and to understand their relationship to God, through parental imagery (e.g., LaHaye 1977:69; Swindoll 1991:95). Consequently, these commentators conclude that while parents must teach their children by example that God is kind and loving, it is equally important to communicate by example the swift and uncompromising nature of divine punishment (Christenson 1970:99-100; Dobson 1976:171-72; Fugate 1980:41-42).

A crucial part of the Conservative Protestant rationale for firm child discipline involves the imperative of conversion (Ellison & Sherkat 1993a; Greven 1977). If the sinful tendencies of youngsters go unchecked, their very souls are believed to be in jeopardy. Many Conservative Protestants fear that if children have not been trained to submit to worldly authority figures, they will also be unable or unwilling to submit themselves to God's supreme authority and guidance, a key requisite of salvation (Daugherty 1991:72; Fugate 1980:190; LaHaye 1977:132). These eternal implications of child discipline are underscored vividly by Dobson (1976:172-73):

If a little child is taught to disrespect the authority of his parents, systematically from the tender years of childhood — to mock their leadership, to "sass" them and disobey their instructions, to exercise extreme self-will from the earliest moments of awareness — then it is most unlikely that this same child will turn his face up to God, about twenty years later, and say humbly, "Here I am, Lord; send me!" To repeat, *a child learns to yield to the authority of God by first learning to submit (rather than bargain) to the leadership of his parents.* (emphasis added)

Given such distinctive and strongly held views on sin and its consequences, religious conservatives stress the urgent need for parents to begin "shaping the

wills" of their children at an early age. Virtually all Conservative Protestant writers agree that Christian parents are morally obligated to use physical "chastisement" with the "rod" because this is the primary biblically ordained response to overt challenges to parental authority (Daugherty 1991:67-68; LaHaye 1977:145-47; Lessin 1979:24; Swindoll 1991:85-88). In support of this claim, they refer to numerous scriptural passages that they interpret as encouraging the use of physical force to shape the will of children (2 Samuel 7:14; Proverbs 23:13-14, 29:15; Hebrews 12:6-8; see Greven 1990:52-55 for an alternative reading of these passages).

Although there is broad consensus among Conservative Protestant parenting writers that corporal punishment is a necessary and important part of child rearing, it bears mentioning that there is less agreement on other issues. For instance, a few writers countenance the expansive use of corporal punishment for virtually any offense, reasoning that most disciplinary infractions ultimately constitute disobedience of one sort or another (e.g., Fabrizio 1969), and some decline to rule out the use of physical punishment to discipline children of almost any age, including adolescents (e.g., Fugate 1980). However, the most prominent conservative religious writers (Dobson 1976, 1987; Swindoll 1991) advocate corporal punishment primarily if not exclusively to confront willful defiance, mainly by toddlers. Dobson (1976, 1987) also suggests that the frequency of physical punishments should diminish as children grow older and develop appropriate internal behavioral controls, and he explicitly discourages parents from spanking or slapping adolescents.

Many Conservative Protestant writers also endorse nonphysical disciplinary strategies, including positive reinforcement, to build a child's self-esteem and maximize compliance (Dobson 1976:78; LaHaye 1977:141-42; Swindoll 1991:104-5), natural or logical consequences as punishment for childish irresponsibility (e.g., Dobson 1976:32), and other techniques.

The foregoing review of popular Conservative Protestant child-rearing literature suggests several guiding hypotheses regarding religious differences in the physical punishment of children. First, because conservative church communities are likely to encourage and support the use of "traditional" parenting practices by their members, we expect that members of Conservative Protestant denominations will use corporal punishment more often than other American parents. Second, given the arguments developed above, we expect that theological conservatism — especially belief that the Bible is inerrant and should guide all human affairs — will be positively associated with the frequency of corporal punishment and that the estimated net effects of denominational affiliation will be mediated via theological conservatism.

Third, given the concern over human sinfulness among many religious conservatives, it is possible that the links between Conservative Protestantism and physical punishment will be partly mediated via parents' assessments of child behavioral patterns. Fourth, in light of the focus on intergenerational hierarchy in conservative religious child-rearing manuals, we expect that the links between indicators of Conservative Protestantism (both affiliation and theological conservatism) and corporal punishment will be partly mediated via parental values that emphasize children's obedience. The remainder of this article designs and executes an empirical test of these hypotheses.

Data and Measures

DATA

To explore the relationships between Conservative Protestantism and the parental use of corporal punishment to discipline toddlers and preschoolers (ages 1-4) and grade school-aged children (ages 5-11), we analyze data from the NSFH. Directed by sociologists at the University of Wisconsin-Madison during 1987-88 (for details, see Sweet, Bumpass & Call 1988), the NSFH is a cross-sectional national probability sample of 13,017 men and women 19 years old and over. This includes oversamples of blacks, Puerto Ricans and Mexican Americans, single-parent families, families with stepchildren, cohabiting couples, and recently married persons.

The NSFH data are ideal for the purposes at hand, and superior to the data used in most prior studies, for several reasons. First, findings from a large national probability sample are more reliable and generalizable than findings from smaller community samples or from convenience samples. Second, the NSFH data set includes information on the Conservative Protestant denominational ties and scriptural beliefs of primary parental respondents, along with extensive information on other parental and household characteristics. Third, the NSFH instrument includes a series of questions about focal children who were selected at random from the roster of household members. By using data on these focal children, we are able to investigate the frequency with which corporal punishment is used to discipline particular youngsters. Further, the NSFH permits us to consider *weekly* frequencies rather than the less precise *annual* or *lifetime* measures that are sometimes used in research on this topic.

Moreover, we are also able to consider the effects of various child characteristics and behaviors on the frequency of corporal punishment. Our analyses are restricted to the primary responding parents of focal toddlers and preschoolers and focal grade school-aged children. These analyses are presented sequentially. Data are weighted to account for the differential probability of selection in both the sample and the household, as well as biases in response rates.

DEPENDENT VARIABLE: FREQUENCY OF CORPORAL PUNISHMENT

The parents of focal children were asked the following questions: "Sometimes children behave well and sometimes they don't. Have you had to spank [FOCAL CHILD] when [he/she] behaved badly in the past week?" Those who responded affirmatively were then asked "About how many times have you had to spank [FOCAL CHILD] in the past week?" Consistent with the findings of previous research (e.g., Wauchope & Straus 1990), corporal punishment is employed more frequently on toddlers and preschoolers than on older children. Roughly 52% of the parents of focal children aged 1-4 reported using corporal punishment during the week prior to the interview, with 20% spanking their child once. To reduce skewness, we truncate this variable so that the highest response category is "six or more" spankings per week; approximately 6% of the parental respondents reported using corporal punishment this frequently. In the analyses that follow, we use ordinary least squares (OLS) regression

techniques to estimate the effects of religious variables and covariates on this outcome measure.

Among the parents of grade school-aged children, some 23% of the parents reported using corporal punishment during the week prior to the NSFH interview. Most of these parents (18% of the total) spanked once. Given the radically skewed distribution of these responses, we construct a dichotomous variable to distinguish those parents who used corporal punishment from those who did not. Because the limitations of OLS regression techniques for modeling dichotomous outcomes are well known, we use logistic regression, in which the dependent variable is the natural log of the odds of a given response category (e.g., spanking versus not spanking) (Aldrich & Nelson 1984).

KEY INDEPENDENT VARIABLES: CONSERVATIVE PROTESTANTISM

A dummy variable is used to identify persons affiliated with Conservative Protestant churches, including the following groups: Southern Baptist, Independent Baptist, other fundamentalist Baptist (Primitive, Foursquare Gospel, etc.), Church of Christ, Church of God, Independent or Open Bible churches, Adventist, Alliance Church, Church of God in Christ, Assemblies of God, Pentecostal, Holiness, Apostolic, and other fundamentalist or evangelical churches.[4]

Theological Conservatism

To measure the types of beliefs regarding religious scripture that were discussed earlier, we use a two-item index (r = .73, p < .001, Cronbach's alpha=.85) based on respondents' agreement with the following statements: (1) "The Bible is God's word and everything happened or will happen exactly as it says." (2) "The Bible is the answer to all important human problems." Responses to each item range from (1) "strongly disagree" to (5) "strongly agree," and the mean score is used as our indicator of *theological conservatism.*[5]

CONTROL VARIABLES

Child Behaviors

Our analyses include several variables tapping parental reports of the frequency with which the focal child exhibited various types of behaviors, moods, and demeanors during the three months prior to the NSFH interview. Primary responding parents of focal children aged 1-4 were read a list of ten such behaviors, including (1) "Is *fussy* or irritable"; (2) "*Loses temper* easily"; (3) "*Bullies,* or is cruel or mean to others"; and (4) *Obeys* or "does what you ask." For each statement, response categories were (1) "not true," (2) "sometimes true," and (3) "often true."[6] Primary responding parents of older children were read a similar list of behaviors (with identical response categories), except that "fussy or irritable" was not included on their list. In preliminary analyses, we explored the association between the other child behaviors on these lists (e.g., "feels sad or depressed," "feels fearful or anxious," "gets along well with others," "tries new things") and the frequency with which children are

subjected to corporal punishment. We detected no relationships between these additional child behaviors and the frequency of corporal punishment.

Authority-Minded Parental Values

To gauge parents' valuation of obedience by their children, we used a two-item index. As part of a larger battery of traits, parents were asked how important it is for their children to "always follow family rules" and to "always do what you ask." Response categories range from (1) "not at all important" to (7) "extremely important." We note that these parental values items differ somewhat from those used in many previous studies (Alwin 1984; Ellison & Sherkat 1993b), in which respondents were asked to rank the importance of a series of child characteristics. The two items identified above are combined into an index ($r = .55$, $p < .001$, Cronbach's $\alpha = .71$), and the mean score is our indicator of *authority-minded parental values.*

Respondent and Household Characteristics

We can have confidence in our findings on the relationships between Conservative Protestantism and the use of corporal punishment by primary parental respondents only when we have controlled for the potentially confounding effects of a range of sociodemographic and background factors, including the following: *age* of respondent (in years); gender of respondent (1 = *female*); race/ethnicity of parent (1 = *black*, 1 = *Hispanic*, 0 = white/Anglo); total *household income* (in tens of thousands of 1986 dollars, logged); parent's *education* (in years); number of *children younger than 5* in the household (besides the focal preschooler); number of *children aged 5 and older* in the household (besides the focal child); and the primary parental respondent's marital status (1 = *single* parent).[7]

Child Characteristics

In addition, we also include dummy variables to identify the child's sex (1 = *female* child), as well as to indicate whether the focal child is a stepchild (1 = *stepchild*) or an adopted child (1 = *adopted*). In the analyses of focal toddlers and preschoolers, we use dummy variables to identify the child's age (1 = *age* 4, 1 = *age 3*, 1 = *age 1*, 0 = *age* 2), because preliminary analyses indicated that the estimated effects of age on the frequency of corporal punishment is curvilinear for children of this age group. We also investigated the association between age and corporal punishment for older children and found the relationship to be generally linear. Therefore we include a linear *age* effect in the models for grade school-aged children. Means and standard deviations of the variables used in the final analyses are displayed in Table 1.

A NOTE ON MISSING DATA

As a rule, valid sample means are substituted for missing values on most predictor variables in order to maximize the effective sample size. Ancillary analyses show that this mean substitution procedure does not significantly alter our results. Because a relatively large proportion of NSFH respondents

TABLE 1: Descriptive Statistics on All Variables[a]

	Children Aged 1-4	Children Aged 5-11
Dependent variables		
Frequency of corporal punishment (continuous)	1.22 (1.67)	—
Use of corporal punishment (dichotomous)	—	.23 (.42)
Religious factors		
Conservative Protestant affiliation	.25 (.43)	.27 (.44)
Theological conservatism	3.36 (1.08)	3.41 (1.10)
Respondent and household characteristics		
Female	.60 (.49)	.55 (.50)
Age	30.10 (6.75)	34.96 (6.71)
Black	.11 (.32)	.14 (.35)
Hispanic	.10 (.30)	.11 (.31)
Single parent	.18 (.39)	.18 (.38)
Education	13.23 (2.79)	12.97 (2.79)
Household income (logged)	.90 (1.38)	1.02 (1.22)
Children younger than 5	.40 (.57)	.29 (.54)
Children aged 5-18	.49 (.81)	.88 (.93)

(approximately 16%) were missing valid information on household income, a dummy variable was included in preliminary models to identify respondents who were initially missing income data. When this variable was consistently unrelated to the frequency of corporal punishment of children of either age group, it was dropped from subsequent models.

TABLE 1: Descriptive Statistics on All Variables[a] (Continued)

	Children Aged 1-4	Children Aged 5-11
Child characteristics		
Female	.46	.51
	(.50)	(.50)
Age	2.47	7.81
	(1.09)	(1.95)
Stepchild	.01	.04
	(.10)	(.21)
Adopted	.01	.03
	(.12)	(.18)
Child behaviors		
Obeys	2.31	2.48
	(.54)	(.53)
Loses temper	1.99	1.86
	(.70)	(.72)
Bullies others	1.36	1.32
	(.57)	(.55)
Fussy	1.98	—
	(.58)	
Child-rearing values		
Authority-minded parental values	5.64	5.91
	(1.04)	(.89)
N	1,393	1,829

[a] Mean descriptive statistics. Standard deviations are in parentheses.

Findings

THE CORPORAL PUNISHMENT OF CHILDREN AGES 1-4

Table 2 presents the results of a series of OLS regression models estimating the net effects of Conservative Protestant denominational ties and conservative Bible beliefs on the frequency with which parental respondents physically punish focal children 1-4 years old. The models are organized in hierarchical fashion; they are designed to take into account (1) the possible confounding effects of various sociodemographic and household background factors and (2) the possible mediating effects of focal child characteristics and parental child-rearing values. Model 1 includes Conservative Protestant religious affiliation, along with controls for sociodemographic characteristics of the household, focal child, and

TABLE 2: Estimated Net Effects of Conservative Protestantism and Covariates on Frequency of Corporal Punishment of Children Aged 1-4[a]

	(1)	(2)	(3)	(4)
Religious factors				
Conservative Protestant affiliation	.229* (.059)	.043 (.011)	.023 (.006)	-.001 (-.001)
Theological conservatism	—	.229*** (.148)	.245*** (.158)	.221*** (.142)
Respondent and household characteristics				
Female	.381*** (.112)	.375** (.110)	.381*** (.112)	.406*** (.120)
Age	-.048*** (-.194)	-.046*** (-.187)	-.032*** (-.129)	-.031*** (-.125)
Black	-.019 (-.004)	-.065 (-.012)	.080 (.015)	.050 (.010)
Hispanic	-.482** (-.087)	-.577*** (-.104)	-.398** (-.072)	-.426** (-.077)
Single parent	-.148 (-.034)	-.138 (-.032)	-.162 (-.037)	-.149 (-.034)
Education	-.031† (-.051)	-.014 (-.024)	-.013 (-.022)	-.008 (-.013)
Household income (logged)	-.040 (-.034)	-.044 (-.036)	-.041 (-.034)	-.045 (-.037)
Children younger than 5	.250*** (.086)	.226** (.078)	.162* (.056)	.154* (.053)
Children aged 5-18	.058 (.028)	.038 (.018)	.006 (.003)	-.003 (-.001)

primary parental respondent. Theological conservatism is added in model 2. In model 3, indicators of positive and negative child behaviors are added, and authority-minded child-rearing values are included in the final model (model 4).

Several findings are especially important. In model 1, Conservative Protestant denominational affiliation emerges as a weak ($p < .06$) positive predictor of the frequency of corporal punishment, even net of the potentially confounding effects of various household, respondent, and child characteristics. As expected, this estimated net effect is completely attenuated with the inclusion of theological conservatism in model 2. It is noteworthy (1) that the estimated net effect of theological conservatism is greater than that of Conservative Protestantism, and (2) that the addition of theological conservatism significantly enhances the predictive power of the model. These empirical patterns indicate that while

TABLE 2: Estimated Net Effects of Conservative Protestantism and Covariates on Frequency of Corporal Punishment of Children Ages 1-4[a]

	(1)	(2)	(3)	(4)
Child characteristics				
Female	-.345***	-.330***	-.262**	-.258**
	(-.103)	(-.099)	(-.078)	(-.077)
Age 1	-.418***	-.432***	-.329**	-.325**
	(-.106)	(-.109)	(-.083)	(-.082)
Age 3	-.237*	-.254*	-.224*	-.231*
	(-.060)	(-.065)	(-.057)	(-.059)
Age 4	-.538***	-.539***	-.483***	-.493***
	(-.137)	(-.137)	(-.123)	(-.125)
Stepchild	.043	.040	.065	.009
	(.002)	(.002)	(.004)	(.001)
Adopted child	.569	.508	.391	.417
	(.040)	(.036)	(.028)	(.030)
Child behaviors				
Obeys	—	—	.266***	-.275***
			(-.086)	(-.089)
Loses temper	—	—	.314***	.317***
			(.133)	(.134)
Bullies others	—	—	.327***	.341***
			(.111)	(.116)
Fussy	—	—	.279***	.278***
			(.097)	(.097)
Child-rearing values				
Authority-minded parental values	—	—	—	.133**
				(.083)
Intercept	3.197	2.244	.669	-.092
Adjusted R^2	.101	.117	.187	.193
(N=1,393)				

[a] OLS regression estimates, metric coefficients. Standardized betas are in parentheses.

† $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

members of Conservative Protestant groups embrace disproportionately conservative views about the Bible and related issues, these beliefs are also held to a lesser degree by members of other religious communities.

Theological conservatism remains a significant predictor in models 3 and 4 with minimal change in the size of the parameter estimate, despite the addition of statistical controls for child behaviors and parental child-rearing values. Thus, it appears that despite their sensitivity to issues of sin and punishment, Conservative Protestant parents are no more likely than other parents to evaluate the behavior of their children in *globally* negative terms — as disobedient, temperamental, and so forth. Further, although parents' theological conservatism is positively associated with valuation of obedience and self-control in children ($r = .34$, $p < .001$), the addition of controls for such authority-minded parental values reduces the estimated net effect of theological conservatism only slightly.[8]

As is frequently the case with survey research, the size of the parameter estimates for religious variables is not overwhelming, and their unique contribution to the overall explanatory power of the models appears relatively modest, approximately 2-3%. However, given the widespread concern about the social antecedents and consequences of corporal punishment among academic and popular experts, our results may have important implications. According to the estimated full model (model 4), each increment in the theological conservatism index of the respondent is associated with an increase of .221 in weekly spankings administered to the focal child aged 1-4. On average, then, a parent with the maximum score of 5 on the theological conservatism index — that is, one who strongly agrees that the Bible is inerrant and contains answers to all human problems — spanks or slaps his or her toddler or preschooler .884 more times per week on average than a theologically liberal parent who receives a score of 1 and strongly disagrees with these scriptural views. To be sure, it is important not to reify these parameter estimates. Nevertheless, assuming a generally constant average rate of physical punishment by parents, these estimates translate into a difference of nearly 50 episodes (.221 x [5-1] x 52 = 45.698) of physical force over the course of a 12-month (52-week) period.[9]

In addition to examining the main effects of religious variables, we also included various cross-product (interaction) terms in model 4 of Table 2 to investigate several contingent relationships. For instance, we examined the possibility that the estimated net effect of theological conservatism on the frequency of corporal punishment is stronger for blacks and Hispanics than for whites/Anglos, for fathers (because many Conservative Protestant households are organized in patriarchal fashion), and for parents with less education and income. In addition, we tested for variations in the effect of theological conservatism by the age and gender of the young child.

We also considered two additional possibilities: (1) It is reasonable to anticipate that, while many parents of various religious backgrounds value obedience in their children, the connection between such authority-minded values and the frequency of corporal punishment is stronger for theological conservatives. (2) Although many parents of all theological stripes may respond to repeated child misbehavior with corporal punishment, it is conceivable that certain types of persistent misbehavior (e.g., disobedience, cruelty to others)

may be especially likely to elicit physical punishment from theologically conservative parents. Ancillary analyses (not shown) yielded no clear support for any of these contingent hypotheses.

THE CORPORAL PUNISHMENT OF CHILDREN AGES 5-11

Table 3 presents logistic regression models estimating the net effects of Conservative Protestantism and covariates on the log odds of corporal punishment by the parents of children 5-11 years old. These models are organized in hierarchical fashion, similar to those presented in Table 2. Again, several patterns warrant attention. Model 1 indicates that membership in a Conservative Protestant denomination is associated with an increase of nearly 50% in the odds that responding parents spanked or slapped their grade school-aged child during the week prior to the NSFH interview, net of the confounding effects of respondent, household, and child characteristics. The inclusion of theological conservatism in model 2 reduces this association but does not attenuate it entirely. As in Table 2, theological conservatism remains a significant predictor of corporal punishment in models 3 and 4. In fact, these models reveal a slight suppressor pattern; the estimated net effect of theological conservatism actually increases somewhat once parental assessments of child behaviors are taken into account. In the full model (model 4), each increment in the theological conservatism index is associated with an increase of nearly 20% in the odds of spanking or slapping a grade school-aged child.

Thus, according to these estimates, the log odds of using corporal punishment in a given week are nearly 75% (.185 x [5-1] = .740) higher for a parent with the maximum score of 5 on this index — i.e., a strong theological conservative — than for a parent of similar background characteristics with the minimum score of 1 on the index — i.e., a strong theological liberal.

Parental affiliation with a Conservative Protestant denomination also bears a weak (and marginally significant) but positive association with corporal punishment throughout this panel of models. In the full model, the log odds of using corporal punishment are roughly 33% (exp [.284]) greater for a parent who reports affiliation with a Conservative Protestant group than for a parent with similar background characteristics who is not affiliated with such a denomination ($p < .08$). As in the analyses of the parents of focal toddlers and preschoolers, tests designed to identify various interactive or contingent relationships with the Conservative Protestant variables turned up no reliable evidence of such patterns.

ESTIMATED NET EFFECTS OF COVARIATES

Although the estimated effects of the covariates themselves are not the primary focus of our study, the NSFH data yield a number of useful insights into the antecedents of corporal punishment. Given the dearth of evidence on this topic, several empirical patterns involving covariates also merit brief discussion. Mothers of very young children and grade school-aged children use corporal punishment more frequently than fathers. Younger parents spank or slap their children more often than older parents, although the estimated net effect of

TABLE 3: Estimated Net Effects of Conservative Protestantism and Covariates on Log Odds of Corporal Punishment of Children Aged 5-11[a]

	(1)	(2)	(3)	(4)
Religious factors				
Conservative Protestant affiliation	.402**	.296*	.288†	.284†
	(1.495)	(1.344)	(1.334)	(1.328)
Theological conservatism	—	.141*	.200***	.185***
		(1.151)	(1.221)	(1.203)
Respondent and household characteristics				
Female	.656***	.623***	.668***	.685***
	(1.927)	(1.865)	(1.951)	(1.984)
Age	-.022*	-.021*	-.011	-.011
	(.978)	(.979)	(.989)	(.989)
Black	.060	.032	.310	.273
	(1.062)	(1.033)	(1.363)	(1.313)
Hispanic	-.044	-.076	.198	.170
	(.957)	(.927)	(1.219)	(1.186)
Single parent	.064	.114	-.002	.005
	(1.066)	(1.120)	(.998)	(1.005)
Education	-.054*	-.048†	-.037	-.032
	(.947)	(.953)	(.964)	(.969)
Household income (logged)	-.030	-.023	-.026	-.025
	(.970)	(.977)	(.975)	(.976)
Children younger than 5	.417***	.413***	.461***	.465***
	(1.517)	(1.511)	(1.585)	(1.592)
Children ages 5-18	.009	-.001	-.036	-.040
	(1.009)	(.999)	(.965)	(.961)
Child characteristics				
Female	-.591***	-.597***	-.544***	-.533***
	(.554)	(.551)	(.580)	(.587)
Age	-.262***	-.266***	-.308***	-.305***
	(.770)	(.767)	(.735)	(.737)
Stepchild	-.069	-.054	-.135	-.128
	(.933)	(.948)	(.874)	(.880)
Adopted child	.424**	.400*	.421*	.413*
	(1.528)	(1.492)	(1.524)	(1.511)

TABLE 3: Estimated Net Effects of Conservative Protestantism and Covariates on Log Odds of Corporal Punishment of Children Aged 5-11[a] (Continued)

	(1)	(2)	(3)	(4)
Child behaviors				
Obeys	—	—	-.845***	-.867***
			(.430)	(.420)
Loses temper	—	—	.585***	.586***
			(1.795)	(1.797)
Bullies others	—	—	.371***	.379***
			(1.450)	(1.461)
Child-rearing values				
Authority-minded parental values	—	—	—	.132†
				(1.141)
Intercept	1.825	1.298	1.229	.412
Model χ^2	204.74	210.05	353.14	356.17
Degrees of freedom	14	15	18	19
Dependent variable mean	.228	.228	.228	.228
Pseudo R^2	.100	.103	.162	.163
(N=1,829)				

[a] Logistic regression coefficients. Exponentiated coefficients are in parentheses.

† $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

parent's age is attenuated in the models for grade school-aged children by controls for parents' assessments of child behavior. Parents with multiple toddlers and preschoolers also physically punish their children — including their older children — more frequently than other respondents. Taken together, these patterns are broadly consistent with the notion that parents who are inexperienced and stressed tend to spank or slap their children more often than others.

Given the substantial body of prior work associating socioeconomic differences in child rearing with *attitudes* toward corporal punishment mentioned at the outset of this study, it is interesting that household income and parental education are only relatively weak predictors of the *use* of corporal punishment, for younger and for older children.[10] While observers have long suggested that African Americans resort to corporal punishment more frequently than whites/Anglos (e.g., Alvy 1987), we find no substantial black-white

differences in the spanking of either toddlers or preadolescents once the effects of relevant sociodemographic and other background factors are held constant. Further, although we find no differences in the corporal punishment of grade school-aged children between Hispanics and whites/Anglos, Hispanic parents report spanking or slapping their toddlers and preschoolers less frequently than white/Anglo parents.[11]

Consistent with previous research (e.g., Wauchope & Straus 1990), we find that child characteristics and parental assessments of child behavior also have important implications for child discipline. In the NSFH, parents' use of physical force peaks around age 2, declines slightly at age 3, and then tails off considerably by age 4. Corporal punishment seems to plateau around ages 5-6, when the child is entering the school-age period, and then declines monotonically thereafter. There is also a substantial gender difference in the use of corporal punishment: Boys of all ages are spanked or slapped much more often than girls. Further, adopted grade school-aged focal children confront elevated risks of physical punishment; they are more than 50% more likely (exp [.413]) than other children to have been spanked or slapped by parental respondents during the week prior to the interview.

It is not surprising that respondents who evaluate their children as obedient use corporal punishment infrequently, while respondents who indicated that their young children engage in undesirable behaviors — i.e., that they are fussy or temperamental, or that they bully others — tend to employ physical force much more often. Given the cross-sectional nature of the NSFH data and the fact that child behaviors are measured from parental assessments rather than from observation, the causal process that underlies these empirical associations remains unclear. It is somewhat surprising that authority-minded parental values (i.e., valuation of children's obedience and conformity) are only modestly associated with the corporal punishment of children ages 1-4 and very weakly linked with the corporal punishment of older children, once the confounding effects of background factors are statistically controlled. This finding highlights the need to explore the factors that mediate the links between abstract, global parental values and concrete parental behaviors (Holden & Edwards 1989).

Discussion

After a long period of scholarly neglect, a small but growing literature has called attention to the relationships between Conservative Protestantism and attitudes toward the corporal punishment of children. Several studies have suggested that members of Conservative Protestant denominations are more likely to support the principle of corporal punishment than other persons (Erlanger 1974; Wiehe 1990). More recently, researchers have shown that such denominational variations may actually reflect the influence of specific conservative theological beliefs, particularly commitment to the doctrine of biblical inerrancy (Ellison & Sherkat 1993a; Grasmick, Bursik & Kimpel 1991).

We have extended this literature by linking conservative theological orientations with the actual *use* of physical punishment to discipline toddlers and preschoolers (ages 1-4) and grade school-aged children (ages 5-11). Analyzing data from a large national survey, we have presented new evidence

that parents who believe (1) that the Bible is the inerrant Word of God and (2) that it provides answers to all human affairs and problems use corporal punishment to discipline their children more frequently than parents with less conservative theological views. In addition, consistent with the thrust of some earlier research (e.g., Ellison & Sherkat 1993a), statistical controls for theological conservatism virtually eliminate the estimated effects of Conservative Protestant denominational affiliation. Further, the link between theological beliefs and the use of physical punishment persists even when the potentially confounding effects of numerous parental and household characteristics, child characteristics and behaviors, and global parental child-rearing values are taken into account.

While our findings underscore the importance of religious values — particularly those associated with Conservative Protestantism — for the study of child discipline, they also raise a number of important questions that warrant clarification in the future. First, the NSFH — like most large-scale surveys — does not contain detailed information on the *type* and *intensity* of corporal punishment. Anecdotal evidence suggests that terms like "spanking" and "slapping" may cover a variety of disciplinary practices, ranging from mild taps or swats on the buttocks to more determined beatings with belts and other objects, to the most severe forms of abuse. Obviously the levels of force and intensity implied by these differences, and the probable consequences for children, may vary considerably. Although some accounts suggest that Conservative Protestants are prone to use particularly violent forms of physical punishment (Capps 1992; Maurer 1982), much more information is needed on the diverse forms of corporal punishment administered by Conservative Protestants and other parents (Bartkowski 1995).

Second, given our finding that the comparatively frequent use of corporal punishment by conservative religious parents is *not* attributable to more negative *global* assessments of children's behavior, we need more information on how parents arrive at decisions on disciplinary responses to *specific* child misbehaviors under *particular* circumstances, and especially on how Conservative Protestant parents may differ from others in their decision-making processes. For instance, given the importance of themes of sin and punishment in Conservative Protestant theology and child-rearing philosophy, parents may be especially prone to interpret specific child misbehaviors as instances of "willful disobedience," to downplay alternative explanations for these misbehaviors (e.g., "childish irresponsibility"), and to deemphasize the role of other types of situational information (e.g., child illness, fatigue or emotional upset, parental stress). In this way, the distinctive theological views of Conservative Protestant parents may be introduced into the decision-making process, possibly leading parents to engage in a specific version of "the fundamental attribution error" — that is, to attribute misbehavior *dispositionally*, as a consequence of the child's character, rather than *situationally*, as a consequence of the circumstances surrounding the event (Grasmick & McGill 1994). A multimethod approach — involving not only standard survey instruments, but also personalized vignettes, in-depth interviews, and perhaps observational methods — will ultimately be required to investigate these and other possible religious variations in the parental decision making process.

Third, our findings almost certainly understate the frequency with which focal children in the NSFH are physically punished. This is the case because our analyses focus solely on the antecedents of spanking or slapping by one respondent parent. However, other adults may also employ physical punishment to discipline the child. This may be especially important for religious conservatives. Because rates of religious intermarriage are relatively low among Conservative Protestants (McCutcheon 1988), one suspects that many of the NSFH focal children — especially toddlers and preschoolers — who are spanked or slapped by a theologically conservative parent may also experience corporal punishment from the respondent's spouse or partner. It is reasonable to hypothesize that homogamous religious conservatives may use corporal punishment more often in part because pastors, friends, and family members are likely to provide sympathetic feedback about the value of "traditional" child-rearing practices. Although an empirical exploration of this issue lies beyond the scope of our study, it will be important for future researchers to examine the role of religious beliefs in shaping the congruence (or incongruence) of parental disciplinary preferences, as well as the ways in which parents' divergent views are harmonized in practice.

Finally, while sociologists have intermittently explored distinctive aspects of Conservative Protestant child-rearing practices (e.g., Nelsen & Kroliczak 1984; Nunn 1964), there is virtually no solid evidence on the short- or long-term effects of corporal punishment on the children of theologically conservative parents.[12] In light of our findings, such research is clearly needed. As we noted at the beginning of this article, numerous academic and popular experts have expressed alarm about the possible impact of physical punishment on various behavioral and developmental outcomes (see Straus 1994b). However, much of the research on the consequences of corporal punishment — especially mild to moderate forms — suffers from methodological weaknesses. With only a handful of exceptions (e.g., Weiss et al. 1992), work in this topic has relied heavily on cross-sectional (as opposed to longitudinal) data or on retrospective accounts of individuals' experiences of corporal punishment. Small community or convenience samples have also been the norm in this area. In addition, some researchers have also combined indicators of corporal punishment with indicators of tougher disciplinary practices — e.g., severe beatings, threats or assaults with knives and guns — to construct measures of "harsh punishment" or the like.

Although such omnibus measures are often correlated with negative child outcomes, they are of little help in specifying the effects of mild to moderate corporal punishment on children. Further, studies have routinely neglected to control for other aspects of child rearing (e.g., amount of quality time between parent and child, level of parental support and nurturance), which may be related to the use of physical punishment and may account for the empirical association between such punishment and negative child and adolescent outcomes (Simons, Johnson & Conger 1994). Moreover, only a few studies have investigated the potential mediating and moderating effects of various contextual factors (e.g., mode of administration, regularity and promptness, clarity of parental explanation) on the consequences of corporal punishment (Kurz 1991; Larzelere 1986, 1994; Larzelere et al. 1989). For these and other reasons, it may

be premature to draw firm conclusions about the effects of mild to moderate corporal punishment.

Nevertheless, if the concerns of many academic and popular child development specialists are on target, then we might expect to find that children from theologically conservative families — who as a group are more likely to experience frequent corporal punishment than others — exhibit heightened risk of various undesirable child and adolescent outcomes. Yet leading Conservative Protestant commentators adamantly contend that prudent physical discipline is essential for *preventing* such outcomes and for raising healthy, happy, Christian children (Christenson 1970; Daugherty 1991; Dobson 1970, 1976; LaHaye 1977; Swindoll 1991). And while the empirical evidence is not unequivocal, a number of studies over the years have reported that religiosity and religious participation among youths — especially those from conservative backgrounds — are associated with lower risk of substance use and abuse, other forms of delinquent behavior, teen pregnancy, and even academic failure (e.g., Beck, Cole & Hammond 1991; Burkett 1993; Stark 1984).

Why might corporal punishment have fewer harmful (perhaps even some beneficial) consequences for Conservative Protestants than for other youths? We suggest two possible reasons. First, the effects of even frequent spanking or slapping may be less damaging in cultural contexts in which the practice of corporal punishment is widely accepted as legitimate. It is conceivable that the discourse within Conservative Protestant families, churches, and communities — where corporal punishment has been the norm for many generations (Greven 1977) — may mitigate the negative emotions (among children and parents) that sometimes accompany physical punishments in other contexts. Second, Conservative Protestants may implement this punishment differently from other parents. For instance, religious child-rearing manuals often devote considerable attention to the details of administering corporal punishment, instructing parents on when to spank, what type of "rod" to use, how and where to hit the child safely, and so forth (e.g., Dobson 1976 and Lessin 1979; see Bartkowski 1995; Ellison & Bartkowski 1996; Greven 1990). As we noted earlier, while a few religious commentators advocate the broad use of relatively harsh corporal punishment (Fabrizio 1969; Fugate 1980), the most popular of these authors (Christenson 1970; Dobson 1970, 1976; Swindoll 1991) recommend the use of moderate physical force only in clear instances of "willful disobedience." Further, they specifically counsel against spanking when the parent is angry or when the behavioral infraction involves childish irresponsibility or some other well-intentioned mistake. These authors are careful to distinguish moderate corporal punishment from physical abuse, and to warn parents about the potential for escalation (Dobson 1976:73-76; Swindoll 1991:82-85). Moreover, several leading Conservative Protestant commentators also advise parents to initiate a period of intimacy and affection with the child after the punishment (Christenson 1970; Dobson 1976).

In short, such mild to moderate spanking would differ substantially from the frustrated, hostile use of physical force decried by academic and popular parenting specialists alike. Of course, to date there is little empirical information on the precise methods of implementation actually used by theologically conservative parents, or the extent to which their practices in this area parallel

those recommended by Dobson and others. Similarly, it is unclear whether the effects of such methods would be more benign than the negative consequences associated with harsher practices. Given that religious conservatives almost certainly constitute the largest and best-organized constituency advocating the use of corporal punishment, the resolution of these issues should be an important priority of researchers in the future.

Notes

1. Here we use the term *Conservative Protestant* as an umbrella term to refer to fundamentalists and other evangelicals. Although the fundamentalist label is sometimes used in precisely such an inclusive fashion, this label is hotly contested. Many evangelicals reject the fundamentalist label because of its pejorative connotation in the media and in broader society and because of various religious and political differences with fundamentalists.

2. We use these insider documents to clarify the worldview of Conservative Protestants with regard to issues of child discipline. We do not assume that these child-rearing manuals influence the practices of parents, but only that these authors generally reflect and articulate the views and sensibilities of religious conservatives. Because our goal is to identify common themes in this popular literature, we necessarily downplay the heterogeneity on some issues within the conservative camp (Bartkowski 1995; Ellison & Bartkowski 1996). In addition, while our work focuses on contemporary religious conservatives, Greven (1977) has called attention to the considerable continuity between their views on child rearing and those of many early American evangelicals.

3. While most prominent Conservative Protestant child-rearing writers endorse such beliefs in original sin, their beliefs are not unanimous. For instance, Campbell (1989, 1992) argues that many writers in this genre exaggerate the salience of themes of sin and punishment at the expense of a more fundamental message of divine love and mercy.

4. Some researchers have suggested that indicators tapping variations in religiosity or religious participation may be more closely related to family attitudes and practices than denominational ties or theological orientations (e.g., Alwin 1986). Therefore, we also constructed an eight-point ordinal measure of the frequency of church attendance similar to that included in the General Social Surveys. This variable was unrelated to the frequency of corporal punishment.

5. The NSFH interview schedule also contains a third item tapping Conservative Protestant sensibilities. This item asks respondents to indicate their level of (dis)agreement with the following statement: "I regard myself as a religious fundamentalist." This item seems less closely related to the theoretical arguments outlined earlier than the items on Bible beliefs. Further, because the term *fundamentalist* has no clear translation in Spanish, the inclusion of this item would require dropping a small number of Hispanics who completed the interview in Spanish. For these reasons, we omit the "fundamentalist" item from our analyses.

6. These indicators of child behaviors are moderately intercorrelated ($.20 < r < .35$). However, they do not comprise a satisfactory index ($\alpha < .55$) in any combination, and the increment to R^2 is greater when they are added individually as predictors of the frequency of corporal punishment.

7. The numbers of Asian and Pacific Islanders, Native Americans/American Indians, and various other categories were too small to analyze, and these groups were dropped from the models that follow. In addition, the numbers of Puerto Ricans (even with oversampling) and Cuban Americans were still too small for separate analyses of these diverse segments of the Hispanic (Latino) population. Although the NSFH also oversampled cohabitors, there are not enough cohabiting respondents who are parents of focal children to permit meaningful analysis.

8. Given the hierarchical structure of these OLS regression models, one possible interpretation of this finding is that authority-minded parents are prone to evaluate their children's behavior negatively. Because the preoccupation with issues of authority and (dis)obedience may influence the encoding and interpretation of child conduct, and hence the overall assessment of child behaviors during the period prior to the interview, one might expect that the estimated

net effect of parental child-rearing values would be slight once the potentially confounding effects of these assessments were taken into account. However, ancillary models (not shown) cast doubt on this interpretation; when we estimated models that *included* parental values and *excluded* parents' reports of child conduct, the results did not change.

9. We acknowledge some potential for response bias. Theologically liberal respondents may tend to understate their "true" or accurate levels of corporal punishment for fear of inviting disapproval from the interviewer or from their fellows. On the other hand, religious conservatives may be inclined to respond accurately, or even to exaggerate their use of such disciplinary techniques, because their use is approved and encouraged within many Conservative Protestant communities. Of course, all respondents, but especially conservative religious parents, who spank or slap their children may also be inclined to underestimate the frequency of these practices, due to the specter of intervention by local or state social service agencies concerned with child protection. Therefore, we doubt that response bias adequately accounts for the empirical patterns discussed in the text.

10. In additional analyses (Ellison, Bartkowski & Segal 1996), we estimated the effects of Conservative Protestantism and covariates on an ordinal measure of the frequency of corporal punishment (with response categories "never," "seldom," "sometimes," or "very often") asked of all NSFH primary respondents who were parents of children under 12 living in the home. While most findings paralleled those reported in the text, it is interesting that both household income and (particularly) parental education emerged as significant inverse predictors of the frequency of corporal punishment in those analyses. The meaning of this discrepancy is not entirely clear, but one possibility is especially intriguing: Some better-educated and more affluent parent respondents may have provided (what for them may be) socially desirable responses, underestimating their frequency of corporal punishment when presented with vague, ordinal response categories. Such "hedging" may have been more difficult for parental respondents who were asked for a more precise count covering the week prior to the NSFH interview.

11. Because southern and rural residence are related to attitudes toward corporal punishment (e.g., Flynn 1994), we included these variables in preliminary models. When both residential variables were consistently unrelated to the frequency of corporal punishment, they were dropped from subsequent analyses.

12. Nelsen and Kroliczak (1984) and Nunn (1964) present evidence that parents' use of the threat that "God will punish" — perhaps particularly characteristic of Conservative Protestant households — may promote self-blame and obedience in children. However, they report that the impact of these parental "coalitions with God" on the development of internal controls seems to depend on images of God and perhaps other factors as well. Additional research is needed to determine whether physical punishment helps to promote internal controls among Conservative Protestant youths.

References

Aldrich, John, and Forrest Nelson. 1984. *Linear Probability, Logit, and Probit Models*. Sage.

Alvy, Kirby T. 1987. *Black Parenting: Strategies for Training*. Irvington.

Alwin, Duane F. 1984. "Trends in Parental Socialization Values: Detroit, 1958-1983." *American Journal of Sociology* 90:359-82.

____. 1986. "Religion and Parental Child-Rearing Orientations: Evidence for a Catholic-Protestant Convergence." *American Journal of Sociology* 92:412-40.

Ammerman, Nancy T. 1987. *Bible Believers*. Rutgers University Press.

Barnhart, Joe E. 1993. "What's All the Fighting About? Southern Baptists and the Bible." Pp. 124-43 in *Southern Baptists Observed*, edited by Nancy T. Ammerman. University of Tennessee Press.

Bartkowski, John P. 1995. "Spare the Rod . . . or Spare the Child? Divergent Perspectives on Conservative Protestant Child Discipline." *Review of Religious Research*. 37:97-116.

Bartkowski, John P., and Christopher G. Ellison. 1995. "Divergent Models of Childrearing in Popular Manuals: Conservative Protestants vs. the Mainstream Experts." *Sociology of Religion* 56:21-34.

Beck, Scott H., Bettie S. Cole, and Judith A. Hammond. 1991. "Religious Heritage and Premarital Sex: Evidence from a National Sample of Young Adults." *Journal for the Scientific Study of Religion* 30:173-80.

Boone, Kathleen C. 1989. *The Bible Tells Them So: The Discourse of Protestant Fundamentalism.* SUNY Press.

Bronfenbrenner, Urie. 1958. "Socialization and Social Class through Time and Space." Pp. 400-25 in *Readings in Social Psychology*, 3d ed., edited by Eleanor Maccoby, Theodore Newcomb, and Eugene Hartley. Holt, Rinehart & Winston.

Burkett, Steven R. 1993. "Perceived Parents' Religiosity, Friends' Drinking, and Hellfire: A Panel Study of Adolescent Drinking." *Review of Religious Research* 35:134-54.

Campbell, Ross, with Pat Likes. 1989. *Kids Who Follow, Kids Who Don't*. Victor Books.

____. 1992. *How to Really Love Your Child*. Victor Books.

Capps, Donald. 1992. "Religion and Child Abuse: Perfect Together." *Journal for the Scientific Study of Religion* 31:1-14.

Cazenave, Noel A., and Murray A. Straus. 1990. "Race, Class, Network Embeddedness, and Family Violence: A Search for Potent Support Systems." Pp. 321-39 in *Physical Violence in American Families: Risk Factors and Adaptations to Violence in 8,145 Families*, edited by Murray A. Straus and Richard J. Gelles. Transaction.

Christenson, Larry. 1970. *The Christian Family*. Bethany House.

Daugherty, Billy Joe. 1991. *Building Stronger Marriages and Families: Making Your House a Home.* Harrison House.

Dobson, James. 1970. *Dare to Discipline*. Living Books/Tyndale.

____. 1976. *The Strong-Willed Child: Birth through Adolescence*. Living Books/Tyndale.

____. 1987. *Parenting Isn't for Cowards*. Word.

Duvall, Debra, and Alan Booth. 1979. "Social Class, Stress, and Physical Punishment." *International Review of Modern Sociology* 9:103-17.

Eckenrode, John, Molly Laird, and John Doris. 1993. "School Performance and Disciplinary Problems among Abused and Neglected Children." *Developmental Psychology* 29:53-62.

Egeland, Byron, Deborah Jacobvitz, and K. Papatola. 1987. "Intergenerational Continuity of Abuse." Pp. 255-76 in *Child Abuse and Neglect: Biosocial Dimensions*, edited by Richard J. Gelles and J.B. Lancaster. Aldine.

Ellison, Christopher G., and John P. Bartkowski. 1996. "Religion and the Legitimation of Violence: The Case of Conservative Protestantism and Corporal Punishment." In *The Web of Violence: From Interpersonal to Global*, edited by Lester Kurtz and Jennifer Turpin. University of Illinois Press. In press.

Ellison, Christopher G., John P. Bartkowski, and Michelle L. Segal. 1996. "Do Conservative Protestant Parents Spank More? Evidence from the National Survey of Families and Households." *Social Science Quarterly*. In press.

Ellison, Christopher G., and Darren E. Sherkat. 1993a. "Conservative Protestantism and Support for Corporal Punishment." *American Sociological Review* 58:131-44.

____. 1993b. "Obedience and Autonomy: Religion and Childrearing Values Reconsidered." *Journal for the Scientific Study of Religion* 32:313-29.

Erlanger, Howard. 1974. "Social Class and Corporal Punishment in Childrearing: A Reassessment." *American Sociological Review* 39:68-85.

Fabrizio, Pat. 1969. *Children — Fun or Frenzy?* Self-published.

Flynn, Clifton P. 1994. "Regional Differences in Attitudes toward Corporal Punishment." *Journal of Marriage and the Family* 56:315-24.

Fugate, J. Richard. 1980. *What the Bible Says about . . . Child-Training.* Alpha/Omega.

Grasmick, Harold G., Robert J. Bursik, and M'Lou Kimpel. 1991. "Protestant Fundamentalism and Attitudes toward Corporal Punishment of Children." *Violence and Victims* 6:283-97.

Grasmick, Harold G., and Anne L. McGill. 1994. "Religion, Attribution Style, and Punitiveness toward Juvenile Offenders." *Criminology* 32:23-46.

Grasmick, Harold G., Carole S. Morgan, and Maryann B. Kennedy. 1992. "Support for Corporal Punishment in the Schools: A Comparison of the Effects of Socio-economic Status and Religion. *Social Science Quarterly* 73:179-89.

Gray, Ellen. 1988. "The Link between Child Abuse and Juvenile Delinquency: What We Know and Recommendations for Policy and Research." Pp. 109-123 in *Family Abuse and Its Consequences*, edited by Gerald T. Hotaling, David Finkelhor, J.T. Kirkpatrick, and Murray A. Straus. Sage.

Greven, Philip. 1977. *The Protestant Temperament: Patterns of Child-Rearing, Religious Experience, and the Self in Early America.* Knopf.

____. 1990. *Spare the Child: The Religious Roots of Punishment and the Psychological Impact of Physical Abuse.* Knopf.

Holden, George W., and Lee A. Edwards. 1989. "Parental Attitudes toward Childrearing: Instruments, Issues, and Implications." *Psychological Bulletin* 106:29-58.

Hotaling, Gerald T., Murray A. Straus, and A.J. Lincoln. 1990. "Intrafamily Violence and Crime and Violence outside the Family." Pp. 431-70 in *Physical Violence in American Families: Risk Factors and Adaptations to Violence in 8,145 Families*, edited by Murray A. Straus and Richard J. Gelles. Transaction.

Kurz, Demie. 1991. "Corporal Punishment and Adult Use of Violence: A Critique of 'Discipline and Deviance.'" *Social Problems* 38:155-61.

LaHaye, Beverly. 1977. *How to Develop Your Child's Temperament.* Harvest House.

Larzelere, Robert E. 1986. "Moderate Spanking: Model or Deterrent of Children's Aggression in the Family?" *Journal of Family Violence* 1:27-36.

____. 1994. "Should the Use of Corporal Punishment by Parents Be Considered Child Abuse? No." Pp. 204-9 in *Debating Children's Lives: Current Controversies on Children and Adolescents*, edited by Marsha A. Mason and Eileen Gambrill. Sage.

Larzelere, Robert E., Michael Klein, Walter R. Schumm, and Samuel A. Alibrando. 1989. "Relations of Spanking and Other Parenting Characteristics to Self-Esteem and Perceived Fairness of Parental Discipline." *Psychological Reports* 64:1140-42.

Lessin, Roy. 1979. *Spanking: Why When How?* Bethany House.

Lindsell, Harold. 1976. *The Battle for the Bible.* Zondervan.

Maccoby, Eleanor, and John A. Martin. 1983. "Socialization in the Context of the Family: Parent-Child Interaction." Pp. 1-101 in *Handbook of Child Psychology*, edited by Paul Mussen. Wiley.

Maddoux, Marlin. 1992. *What Worries Parents Most: Survival Strategies in a Chaotic World.* Harvest House.

Maurer, Adah. 1974. "Corporal Punishment." *American Psychologist* 29:614-26.

____. 1982. "Religious Values and Child Abuse." Pp. 57-63 in *Institutional Abuse of Children*, edited by Ranae Hanson. Haworth Press.

McCutcheon, Allan L. 1988. "Denominations and Religious Intermarriage: Trends among White Americans in the Twentieth Century." *Review of Religious Research* 30:213-27.

McNamara, Patrick H. 1985. "The New Christian Right's View of the Family and Its Social Science Critics: A Study in Differing Presuppositions." *Journal of Marriage and the Family* 47:449-58.

Nelsen, Hart M., and Alice Kroliczak. 1984. "Parental Use of the Threat 'God Will Punish': Replication and Extension." *Journal for the Scientific Study of Religion* 23:267-77.

Nunn, Clyde Z. 1964. "Child-Control through a 'Coalition with God.'" *Child Development* 35:417-32.

Rose, Susan D. 1988. *Keeping Them Out of the Hands of Satan: Evangelical Schooling in America.* Routledge, Chapman & Hall.

Simons, Ronald L., Jay Beaman, Rand D. Conger, and W. Chao. 1993. "Childhood Experience, Conceptions of Parenting, and Attitudes of Spouse as Determinants of Parental Behavior." *Journal of Marriage and the Family* 55:91-106.

Simons, Ronald L., Christine Johnson, and Rand D. Conger. 1994. "Harsh Corporal Punishment versus Quality of Parental Involvement as an Explanation of Adolescent Maladjustment." *Journal of Marriage and the Family* 56:591-607.

Simons, Ronald L., Les B. Whitbeck, Rand D. Conger, and Wu Chyi-In. 1991. "Intergenerational Transmission of Harsh Parenting." *Developmental Psychology* 27:159-71.

Stark, Rodney. 1984. "Religion and Conformity: Reaffirming a Sociology of Religion." *Sociological Analysis* 45:273-82.

Straus, Murray A. 1991. "Discipline and Deviance: Physical Punishment of Children and Violence and Other Crime in Adulthood." *Social Problems* 38:133-54.

____. 1994a. "Should the Use of Corporal Punishment by Parents Be Considered Child Abuse? Yes." Pp. 197-203 in *Debating Children's Lives: Current Controversies on Children and Adolescents,* edited by Marsha A. Mason and Eileen Gambrill. Sage.

____. 1994b. *Beating the Devil Out of Them: Corporal Punishment in American Families and Its Effects on Children.* Lexington Books.

Straus, Murray A., and Richard J. Gelles. 1986. "Social Change and Change in Family Violence from 1975 to 1985 as Revealed by Two National Surveys." *Journal of Marriage and the Family* 48:465-79.

Sweet, James A., Larry Bumpass, and Vaughn Call. 1988. "The Design and Content of the National Survey of Families and Households." Center for Demography and Ecology, University of Wisconsin, Madison.

Swindoll, Chuck. 1991. *The Strong Family: Growing Wise in Family Life.* Multnomah.

Taylor, Leslie, and Adah Maurer. 1985. *Think Twice: The Medical Effects of Physical Punishment.* Generation Books.

Wald, Kenneth D., Dennis E. Owen, and Samuel S. Hill. 1989. "Habits of the Mind? The Problem of Authority in the New Christian Right." Pp. 93-108 in *Religion and Political Behavior in the United States,* edited by Ted G. Jelen. Praeger.

Wauchope, Barbara A., and Murray A. Straus. 1990. "Physical Punishment and Physical Abuse of American Children: Incidence Rates by Age, Gender, and Occupational Class." Pp. 133-48 in *Physical Violence in American Families: Risk Factors and Adaptations to Violence in 8,145 Families,* edited by Murray A. Straus and Richard J. Gelles. Transaction.

Weiss, Bahr, Kenneth A. Dodge, John E. Bates, and Gregory S. Pettit. 1992. "Some Consequences of Early Harsh Discipline: Child Aggression and a Maladaptive Social Information Processing Style." *Child Development* 63:1321-35.

Widom, C.S. 1989. "Child Abuse, Neglect, and Violent Criminal Behavior." *Criminology* 27:251-71.

Wiehe, Vernon R. 1990. "Religious Influence on Parental Attitudes toward the Use of Corporal Punishment." *Journal of Family Violence* 5:173-86.

Wolfe, David A. 1987. *Child Abuse: Implications for Child Development and Psychopathology.* Sage.

Gender Ideology and Perceptions of the Fairness of the Division of Household Labor: Effects on Marital Quality*

THEODORE N. GREENSTEIN, *North Carolina State University*

Abstract

Under what circumstances will married women perceive inequalities in the division of household labor as unfair? This research develops and tests a model based on relative deprivation theory that suggests that gender ideology functions as a moderator variable in a process through which inequalities in the division of household labor come to be seen as inequities. Using data from the National Survey of Families and Households, three empirical tests of the model provide evidence that inequalities in the division of household labor are more strongly related to perceptions of inequity for egalitarian than for traditional wives, and that perceptions of inequity are more strongly related to perceived quality of the marital relationship for egalitarian than for traditional wives. The findings suggest that researchers studying the division of household labor need to shift their focus away from analyses of objective inequalities and toward the study of perceived inequity.

Gender-based inequality in the division of household labor has become a key issue for scholars of marriage and the family. Research has found that although married women perform the majority of household labor, relatively few wives feel that this arrangement is unfair. The purpose of this study is to present and test a model that explains how and why *inequalities* in the household division of labor come to be perceived by some wives as *inequities,* and how those perceptions affect marital quality.

** The National Survey of Families and Households was funded by a grant (#HD21009) from the Center for Population Research of the National Institute of Child Health and Human Development. The survey was designed and carried out at the Center for Demography and Ecology at the University of Wisconsin-Madison under the direction of Larry Bumpass and James Sweet. The fieldwork was done by the Institute for Survey Research at Temple University. Maxine Atkinson, April Brayfield, Donald Tomaskovic-Devey, and two anonymous* Social Forces *reviewers made valuable comments on an earlier draft of this article. The analyses and interpretations herein are those of the author. Please direct correspondence about this manuscript to Theodore N. Greenstein, Department of Sociology and Anthropology, North Carolina State University, Raleigh, NC 27695-8107 (Internet ted_greenstein@ncsu.edu).*

 Social Forces, March 1996, 74(3):1029-1042

Two key findings in this literature have been that (1) wives perform far more household labor than do their husbands, and that (2) the kinds of household tasks performed by wives and husbands differ. Blair and Lichter's (1991) recent analysis of the National Survey of Families and Households produced results typical of most studies. They found that women perform an average of 33 hours of household tasks (exclusive of child care) per week, compared to men's average of 14 hours. Many researchers (e.g., Blair & Lichter 1991; Brayfield 1992; Mederer 1993) have also noted that household labor remains highly segregated by sex; those tasks that have traditionally been thought of as "women's work" (e.g., cooking, laundry, housecleaning) are primarily performed by women, while "male" tasks such as yard work and auto maintenance are done primarily by men.

The central concern of this article is not with the amount or *content* of household tasks nor is it with the sources or amount of *inequality* in the division of those tasks. Rather, the focus is on the consequences and perceptions of those inequalities. The model starts from the empirical generalization that husbands tend not to share equally in the division of household labor, even when their wives are employed full-time. The consensus of dozens of studies of this issue is that the level of a husband's contribution to family work is not strongly related to his wife's employment status (e.g., Berk 1985; Huber & Spitze 1983; Pleck 1985; Shelton 1990). A finding of Demo and Acock (1993) on this issue was that while husbands of employed wives contribute on average 4.3 more hours per week on chores than do husbands of unemployed wives, the division of household labor remains strikingly unequal: employed wives' proportion of total hours spent on household chores is still about 72% and about 81% for unemployed wives. The division of household labor is relatively traditional — that is, the wife performs a far greater proportion of household tasks than does her husband — in households where the wife earns more than her husband (Atkinson & Boles 1984) and even in households where the husband is not employed (Brayfield 1992). When married women work outside the home, their combination of market and nonmarket work is likely to force them into working what Hochschild calls the second shift (Hochschild with Machung 1989).

Unquestionably, there *are* gender-based inequalities in the division of household labor. However, even though most wives do the bulk of household tasks, relatively few wives feel that this division of labor is unfair (see, e.g., Benin & Agostinelli 1988; Berk 1985; Pleck 1985; Yogev 1981). The key issue here deals with *perceptions* of justice: under what circumstances will *inequalities* in the division of household labor come to be perceived as *inequities*? To answer this question, we need to consider the theoretical literature on justice.

Relative deprivation theory (Crosby 1976) makes several crucial assumptions about the nature of justice. First, it is the *subjective* evaluation of outcomes that determines whether one feels unjustly treated. Second, individuals must *value* an outcome in order for justice phenomena to become relevant. Third, perceptions of injustice or inequity are the results of *comparison processes*. It is axiomatic that no justice phenomenon can arise in the absence of a stable frame of reference. Finally, the theory suggests that there is *preference for equity* in relationships, and that when individuals perceive that they have been unjustly

treated they will feel distressed and will be motivated to take action to restore equity.

From a relative deprivation perspective, an outcome is equitable when it is similar to that of some comparative referent (another person, a group, past experience, one's expectations). When an outcome falls short of what is expected, the individual in question will feel deprived in a relative sense (even though in absolute terms the individual may not be deprived at all). The relative deprivation approach has been quite fruitful in the study of justice in the family (e.g., see Major 1993; Sanchez 1994; Thompson 1991; Thompson & Walker 1989).

Major (1987) argues that three key elements contribute to perceptions of fairness: outcome values, comparison referents, and justifications. Gender ideology provides a mechanism through which these elements may be understood in the context of the perceived fairness of the division of household labor. Traditional women differ from nontraditional or egalitarian women in what they want out of their marital relationship (outcome values), the standards they use to judge their reward structure (comparison referents), and their perception of the legitimacy of reward allocation schemes (justifications).

Crucial to my explanation is the assumption that married women will have different reactions to an unequal division of household labor that are contingent on their gender ideology. I argue that the wife's gender ideology functions as a kind of lens through which inequalities in the division of household labor are viewed. According to the nature of that lens, such inequalities may be perceived as unfair or fair, inequitable or equitable. In effect, gender ideology serves to provide married women with a comparative referent.

Although the effects of gender ideology and inequalities in the household division of labor on marital quality have been studied (for example, Blair & Johnson 1992; Piña & Bengtson 1993), these effects have typically been modeled in an additive rather than interactive fashion. That is, most researchers that have included measures of ideology in their analyses of the household division of labor have assumed that gender ideology functions as a *mediator* variable (but see Bielby & Bielby 1992 and John, Shelton & Luschen 1995 for examples of research that conceptualize ideology as a moderator variable). The model proposed here explicitly assumes that inequalities in the division of household labor affect factors such as marital quality and perceptions of fairness in different ways for traditional than for nontraditional or egalitarian wives. In other words, I conceptualize gender ideology as a *moderator* (rather than a mediator) variable.

Why should married women of different gender ideology differ in their reactions to inequalities in the household division of labor? In the context of Major's (1987) model, I can suggest three possibilities. First, wives probably differ in the relative value they place on the various outcomes and characteristics of their marriages. Traditional wives might value stability and harmony in their relationships, while egalitarian wives might be more concerned with independence and autonomy.

Second, wives probably differ in their choice of comparison referents. A wife might not perceive the inequality in the household division of labor as inequitable because she might feel that her rewards are consistent with either

(1) What she expects to receive based on her past experiences, or
(2) What she expects to receive based on what she perceives that similar wives typically receive, or
(3) The referent group to which she compares herself

Gender ideology probably affects all three of these possible comparison referents. For example, traditional women might base their expectations about the household division of labor on their past experiences or those of their siblings or mothers; they might base them on what they perceive other wives receive; or they might base their expectations on the reference group of women who are ideologically and economically similar to themselves. Egalitarian women, on the other hand, will use very different comparison referents: other egalitarian wives, or perhaps expectations about how egalitarian marriages are supposed to be structured.

Third, wives may differ in their reactions to inequalities in the division of household labor according to whether they see the outcome as the result of a legitimate or justifiable process. If a wife truly believes that married women — even those employed full-time — are "supposed" to do the bulk of the housework with little assistance from their husbands, then she will probably see this inequality as resulting from a legitimate process, and not view it as unjust or inequitable. Some research (e.g., Crosby 1982; Kanter 1977) even implies that wives who are especially economically dependent on their husbands and cannot easily change their situation may come to believe that they "deserve" this inequality rather than concluding that they are being treated inequitably.

When a married woman perceives the division of household labor to be unjust, her reactions may take one of several forms (Major 1987; Mark & Folger 1984). She may direct her reactions to the perceived inequity at the *system* of marriage itself. She might attempt to change the structure of her marriage by encouraging her husband to contribute more household labor; failing that, she might become resentful and unhappy with her marriage, increasing the level of conflict and likelihood of separation and divorce. Alternatively, she might attempt to cognitively restructure the situation in an attempt to resolve the injustice. For example, she might devalue the *outcome* in question ("I don't really care if I have to do all the housework") or her *contributions* ("What I do isn't that important, anyway") or *revalue* the contributions of her partner ("What he does is more important than what I do" or "His job pays more").

Hypotheses

What predictions concerning the perceptions of married women can we base on the relative deprivation model? First, it is reasonable to expect that wives' reactions to the unequal division of household labor will be contingent upon their gender ideology. Traditional women — those who accept the "doctrine of separate spheres" where the ideal marriage consists of an economically successful husband and a dependent, nurturing wife — will view the additional workload created by the unequal division of household labor as consistent with their ideology and therefore not necessarily unjust. Nontraditional or egalitarian women, on the other hand, will find this inequality in the division of household

labor onerous and inequitable. Consequently, I expect to find that *inequalities* in the division of household labor are more likely to be perceived as *inequities* by nontraditional wives than by traditional wives. That is, the extent of the inequality in the division of household labor will be a better predictor of feelings of injustice or unfairness for nontraditional wives than for traditional wives. Gender ideology and the actual division of household labor should interact in their effects on perceived fairness of the division of household labor.

Second, I expect to find that gender ideology also moderates the effects of wives' perception of injustice in the division of household labor on the perceived quality of their marriages. While perceived fairness will probably affect the perceived quality of marriage for all married women, it should have a stronger effect for egalitarian or nontraditional wives than it does for traditional wives. Gender ideology and perceived fairness of the division of household labor should interact in their effects on perceived marital quality.

Data and Measurement

The data for these analyses come from the first wave of the National Survey of Families and Households (NSFH; Sweet, Bumpass & Call 1988). The NSFH is a national probability sample of 13,017 adult respondents interviewed in 1987 or 1988. In married couple and cohabiting households, a questionnaire was also administered to the respondent's spouse or partner. The analyses here focus on the 4,960 married couples (out of an unweighted total of 5,640) where both wife and husband responded to the survey and provided codable responses to all the items under study here. All analyses used the NSFH couple weights to adjust for oversampling, differential probabilities of selection, and differential response rates.

DEPENDENT VARIABLES

Perception of Fairness of the Division of Household Labor

NSFH respondents were asked "How do you feel about the fairness in your relationship in each of the following areas?": household chores, working for pay, spending money, and child care. The responses to the household chores item were used as an indicator of perceived fairness of the division of household labor. Response choices and their codings were "very unfair to me" (coded −1), "somewhat unfair to me" (−0.5), "fair to both" (0), "somewhat unfair to her/him" (+0.5), and "very unfair to her/him" (+1).

Perceived Marital Stability

NSFH respondents were asked "It is always difficult to predict what will happen in a marriage, but realistically, what do you think the chances are that you and your husband/wife will eventually separate or divorce?" Response choices were "very low" (coded as 5), "low" (4), "about even" (3), "high" (2), and "very high" (1).

Marital Happiness

NSFH respondents were asked, "Taking all things together, how would you describe your marriage?" Response choices ranged from "very unhappy" (1) to "very happy" (7).

INDEPENDENT VARIABLES

Indicators of Household Labor

Each spouse was asked to indicate "the approximate number of hours per week that you, your spouse/partner, or others in the household normally spend doing" preparing meals, washing dishes, cleaning house, outdoor tasks, shopping, washing and ironing, paying bills, auto maintenance, and driving other household members to work and school. Time spent in these tasks was totaled into measures of total hours spent by wives, husbands, children, and others doing household labor. An overall measure of the proportion of total household labor performed by the wife was then created.

Inspection of the NSFH household labor data shows that many respondents reported spending more than 168 hours per week on a combination of domestic and market work. Such responses may reflect invalidity in the self-reports of hours spent in household labor or the fact that many household tasks are performed concurrently. To adjust for this problem, any respondent who reported spending more than 100 hours per week in any one domestic work activity was coded as having spent 100 hours in that activity.

Wife's Traditionalism

NSFH respondents were asked a series of questions that were used to construct a summated scale of gender ideology. On a 1-7 scale where 1 = "strongly approve" and 7 = "strongly disapprove," respondents were asked how much they approved of "mothers who work full-time when their youngest child is under age five" and "mothers who work part-time when their youngest child is under age five." On a 1-5 scale, where 1 = "strongly agree" and 5 = "strongly disagree," respondents were also asked how much they agreed with the following four items: "It is much better for everyone if the man earns the main living and the woman takes care of the home and family"; "Preschool children are likely to suffer if their mother is employed"; "Parents should encourage just as much independence from their daughters as in their sons"; and "If a husband and wife both work full-time, they should share housework tasks equally." The six items were standardized and summed, yielding a Cronbach's alpha of .68. The summated scale was standardized to a mean of 100 and standard deviation of 15. Item scoring was reversed to produce a scale in which higher scores indicate more traditional ideologies.

CONTROL VARIABLES

There are a number of background characteristics of the wives, their husbands, and their marriages that I included in the model: race, wife's education in years, wife's and husband's market hours worked per week, census region of residence, residence in an SMSA, marital duration in years, and number of

children in the household. Each of these variables is potentially related to the independent and dependent variables described above.

Analysis and Results

DESCRIPTIVE STATISTICS

Table 1 presents weighted means and standard deviations for the entire sample of women and (for descriptive purposes only) three gender ideology subgroups created by dividing the total sample into thirds based on their gender ideology scale score. Overall, the women in the sample are fairly representative of the population of U.S. women who were married during the study period (1987-88). They had been married an average of 22.9 years. About 7 percent are African American, and about 6 percent are Hispanic. The women tended to have some postsecondary education, with a mean of 12.6 years of formal education. Most of the wives were employed at the time of the interview, averaging about 18.5 hours per week in the paid labor force. The households had an average of 1.03 children under the age of 18 years.

The descriptive statistics on the division of household labor are consistent with those of previous studies. Wives averaged about 37 hours per week in household tasks (exclusive of child care), while husbands averaged about 16 hours. Wives performed about 64% of all household labor (including household labor performed by children and others in the household).

Not surprisingly, the women in the gender ideology subgroups differ substantially (the differences in group means are statistically significant beyond the .05 level for all variables except metropolitan residence and living in the Northeast or North Central regions). For example, egalitarian women were more likely to feel that the division of household labor was unfair, perceived a greater chance of their marriages ending in divorce, and were less likely to respond that they were "Very happy" with their marriages. Traditional women in the sample tend to have been married longer, to have less formal education, to spend more time doing domestic labor, and to work fewer hours outside the home. Consistent with earlier studies, traditional women performed a greater proportion (67%) of the domestic labor than egalitarian women (59.9%).

ANALYSES OF PERCEPTIONS OF FAIRNESS OF THE DIVISION OF HOUSEHOLD LABOR

The second column of Table 2 presents the results of an OLS regression analysis for all wives modeling the effects of 13 predictor variables on perceived fairness of the division of household labor. Overall, this model accounts for about 4.4% of the variation in the dependent variable. Six of the predictors have significant effects on perceptions of fairness of the division of household labor. Wives with longer marital durations are more likely to perceive the division of household labor as fair, as are Hispanic wives. Perception of the division of household labor as fair decreases as the wife's total hours spent performing household tasks and hours in market labor increase, but increases as the husband's contributions to household labor increases. Most important, there is a statistically significant interaction between the wife's share of the domestic labor and

TABLE 1: Descriptive Statistics for Variables in Analyses[a]

	All Wives	Wife's Gender Ideology: Egalitarian	Transitional	Traditional
Unweighted sample size	4,960	1,656	1,647	1,657
Perceived fairness of the division of household labor	-0.14 (0.32)	-0.17 (0.34)	-0.15 (0.32)	-0.11 (0.31)
Perceived marital stability	4.66 (0.69)	4.56 (0.78)	4.67 (0.65)	4.73 (0.65)
Wife's marital happiness	6.08 (1.26)	6.01 (1.29)	6.09 (1.21)	6.13 (1.27)
Duration of marriage in years	22.9 (15.17)	18.2 (13.70)	23.6 (15.12)	26.4 (15.39)
Wife's education (in years)	12.6 (2.8)	13.5 (2.6)	12.3 (2.6)	12.0 (2.8)
Wife's ethnicity				
Black (percent)	6.6	8.2	6.3	5.4
Hispanic (percent)	6.4	4.0	5.7	9.3
Region of residence				
Northeast (percent)	19.7	21.1	19.3	18.9
North central (percent)	26.0	27.4	26.8	24.0
South (percent)	35.3	34.2	34.0	37.6
Percent in SMSA	73.7	75.9	73.0	72.6
Number of children in household	1.03 (1.25)	1.10 (1.14)	0.96 (1.20)	1.04 (1.37)
Hours spent in:				
Domestic labor by wife	36.8 (23.4)	32.4 (20.9)	37.4 (23.5)	40.2 (24.6)
Domestic labor by husband	16.1 (15.4)	17.6 (15.4)	16.1 (16.0)	14.8 (14.7)
Market labor by wife	18.5 (19.8)	26.1 (19.9)	18.2 (19.6)	12.0 (17.5)
Market labor by husband	33.4 (22.2)	38.0 (19.9)	32.2 (22.4)	30.5 (23.3)
Wife's percent of domestic labor (percent)	63.9	59.9	64.4	67.0
Wife's traditionalism	100.3 (15.5)	82.2 (9.1)	100.5 (3.5)	116.0 (8.1)

[a] Table entries are means and, in parentheses, standard deviations (for continuous variables) and percentages for categorical variables.

gender ideology: as gender ideology becomes more traditional, the effect of inequalities in the household division of labor on perceived fairness decreases.

I should point out that, although these findings are consistent with most of the empirical studies of the division of household labor, there are marked differences with Blair and Johnson's (1992) findings from the same NSFH data set. They found no significant effects of wives' market labor on perceptions of unfairness, and noted that "ideological variables appear to explain very little of the variation in wives' perceptions of fairness" (579). There are sizable differences between the samples under analysis in the present study and that of Blair and Johnson; my sample size is 4,960 couples, whereas Blair and Johnson's is 778. Part of this difference is attributable to Blair and Johnson's exclusion of black and Hispanic wives from their analyses, and part is due to the limitation of their sample to primary respondents only.

ANALYSES OF MARITAL HAPPINESS

Column 3 in Table 2 presents the results of an OLS regression analysis for all wives modeling the effects of 13 predictor variables on the measure of the wife's self-reported marital happiness. Overall, the model accounts for about 7.5% of the variation in the dependent variable. In general, the results from this model are consistent with those of similar studies (e.g., Blair 1993; Perry-Jenkins & Folk 1994). African American wives tend to have lower levels of marital happiness; wives in the Northeast and South tend to have higher marital happiness scores. Wives' marital happiness increases as market labor hours increases. The interaction between wife's gender ideology and perceptions of fairness, although not reaching conventional levels of statistical significance (the effect is statistically significant using a one-tailed test at the .07 level), is in the predicted direction: perception of fairness of the division of household labor affects marital happiness more strongly for egalitarian wives than for traditional wives.

ANALYSES OF PERCEPTIONS OF MARITAL STABILITY

Column 4 of Table 2 presents the results of an OLS regression analysis for all wives modeling the effects of 13 predictor variables on the measure of the wife's perceptions of marital stability. Overall, the model accounts for about 9.3% of the variation in the dependent variable. Women with longer marital durations, those with more education, those living in the Northeast and North Central regions, and those employed more hours tend to perceive that their marriages are more stable. On the other hand, being African American and having more children negatively affect perceptions of marital stability. The interaction between perceived fairness of the division of household labor and gender ideology is statistically significant and in the predicted direction: perceived fairness affects perceived marital stability more strongly for egalitarian wives than it does for traditional wives.

TABLE 2: OLS Regressions Predicting Perceived Fairness of Division of Household Labor and Measures of Marital Quality[a]

Effect	Fairness of Division of Household Labor	Wife's Marital Happiness	Perceived Marital Stability
Constant	0.102	6.371*	4.367*
Perceived fairness of division of household labor	—	1.463*	0.941*
Duration of marriage (in years)	0.001*	-0.003	0.006*
Wife's education (in years)	-0.001	-0.001	0.009*
Wife's ethnicity			
Hispanic	0.059*	0.03	-0.051
African American	0.012	-0.258*	-0.196*
Metropolitan residence?	0.003	-0.111*	-0.018
Region of residence			
Northeast	-0.013	0.117*	0.069*
North central	-0.009	0.062	0.055*
South	-0.026*	0.159*	-0.012
Number of children in household	-0.021*	-0.065*	-0.028*
Hours spent in:			
domestic labor by wife	-0.0008*	0.0004	-0.0013
domestic labor by husband	0.0017*	0.0014	-0.0003
market labor by wife	-0.0017*	0.0010*	0.0001*
market labor by husband	-0.0002	-0.0035	-0.002
Wife's percent of domestic labor	-0.446*	-0.046	0.039
Wife's traditionalism	-0.0014	0.00000	0.0017*
Interaction between wife's traditionalism and wife's percent of domestic labor	0.004*	—	—
Interaction between wife's traditionalism and wife's perception of fairness of division of household labor	—	-0.005	-0.006*
Adjusted model R^2	.044*	.075*	.093*

[a] Table entries are unstandardized (metric) regression coefficients. Unweighted sample size for all analyses is 4,960.

* $p < .05$ (coefficient at least twice its standard error)

MULTIVARIATE TEST OF THE INTERACTION HYPOTHESIS

To assess the overall hypothesis that gender ideology and perceptions of fairness of the division of household labor interact in their effects on perceived general marital quality, I performed a multivariate test across the two analyses of marital quality measures reported above. Wilks's lambda for these data is .998 with a corresponding F statistic of 4.60 at 2 and 4,941 df, $p = .011$. Thus, we can confidently reject the null hypothesis that there is no interaction between gender ideology and perceptions of fairness of the division of household labor in their effects on marital quality. Perceptions of fairness of the division of household labor *do* appear to affect perceptions of marital quality more strongly for egalitarian wives than for traditional wives.

Discussion

The central issue under study in this research was, under what circumstances will wives perceive inequalities in the household division of labor as inequitable or unjust? Even though most wives do most of the household labor, relatively few wives feel that their household division of labor is unfair. Apparently, *inequalities* in the division of household labor are not always perceived by married women to be *inequities*. Applying relative deprivation theory to the analysis of this situation, I assumed that gender ideology provides a comparative referent that married women use to evaluate the fairness or justice of their household division of labor.

Given that husbands tend not to share equally in the division of household labor, all women — regardless of gender ideology — have to work more total hours (that is, both market and nonmarket labor) as their participation in the market labor force increases. For women holding traditional gender attitudes, it is plausible that this "second shift," while undoubtedly stressful, would not necessarily contribute to perceptions of marital quality because such unequal allocations of responsibility for household labor are consistent with the expectations of a married woman holding a traditional gender ideology. Consequently, traditionally oriented women would not necessarily perceive these inequalities as inequitable. Nontraditional women, on the other hand, are likely to perceive such inequalities as inequitable; their expectations of an egalitarian relationship with their spouse are violated by being forced to work a "second shift." Their reaction to this inequity would typically manifest itself as increases in conflict with their husbands, a decline in marital satisfaction and happiness, and a greater likelihood that their marriages would experience a separation or divorce.

The first hypothesis predicted that wives who subscribe to nontraditional or egalitarian gender ideologies are most likely to view inequalities in the division of household labor as inequitable. The analyses reported in Table 2 confirm this hypothesis; inequalities in the division of household tasks are related to the likelihood that an egalitarian wife sees the division of household labor as unfair; for traditional wives, the division of household labor is not related to perceptions of injustice.

My second hypothesis suggested that perceptions of inequity in the household division of labor will affect marital quality more strongly for egalitarian or nontraditional wives than it does for traditional wives. Two empirical tests — effects on wives' perceptions of marital stability and wives' reports of marital happiness — confirmed this hypothesis.

Some authors have suggested that ideological factors are not important determinants of wives' perceptions of fairness of the division of household labor. Blair & Johnson (1992), for example, concluded that "ideological variables appear to explain very little of the variation in wives' perceptions of fairness" (579), although they found a statistically significant effect of family labor ideology and perceptions of fairness of the division of household labor for employed wives, particularly for younger wives. The present analyses suggest that the role of gender ideology in this process is that of a moderator variable. While ideology does have effects on perceptions of inequity in the division of household labor — egalitarian wives are more likely to perceive the division of household labor as unfair than are traditional wives — its most theoretically compelling effect is that it moderates the perceptions of inequities as well as perceptions of marital quality.

One important direction for future research is to study the role of husbands' gender ideologies on their perceptions of fairness and marital quality. The present research used relative deprivation theory to understand the linkage between wives' ideologies and perceptions. However, it seems unlikely that relative deprivation theory will prove to be a fruitful approach to understanding husbands' perceptions. The motivating factor underlying justice phenomena is the perception that there is a gap between what one expects to receive and what one anticipates, leading to a belief that one is being inequitably treated. For most husbands, this gap (at least relative to domestic work) simply does not exist; the husbands in this study performed on average only about 30% of the domestic work. Thus it seems unlikely that husbands' perceptions and behaviors can be meaningfully modeled within a relative deprivation framework.

In conclusion, this research finds that relative deprivation theory can be useful in understanding why inequalities in the household division of labor are not necessarily perceived by married women as unfair or inequitable and how this justice phenomenon is related to marital quality. Gender ideology serves as a comparative referent to which married women compare their own division of labor. When their own outcomes are consistent with that comparative referent, the situation is likely to be judged as fair. When these outcomes are inconsistent with that referent, the situation is likely to be perceived as unfair. Consistent with this theory, traditionally oriented wives are less likely to judge their household division of labor as unfair than are nontraditional or egalitarian wives. Further, perceptions of fairness or justice in the household division of labor are more strongly related to marital quality for egalitarian than for traditional wives.

These findings are consistent with recent studies (e.g., Blair 1993; Perry-Jenkins & Folk 1994) that find that wives' *perceptions* of the fairness of the division of household labor are better predictors of marital conflict than the *actual* extent of inequality. Relative deprivation theory suggests that the unequal division of household labor will have consequences for marital quality only

when that inequality is perceived to be unjust. One important implication of these findings is to shift our focus from the *objective* level of inequality in the division of household labor to the *subjective* or cognitive processes by which these inequalities come to be perceived as inequities.

References

Atkinson, Maxine P., and Jacqueline Boles. 1984. "WASP: Wives as Senior Partners." *Journal of Marriage and the Family* 46:861-70.

Benin, Mary Holland, and Joan Agostinelli. 1988. "Husbands' and Wives' Satisfaction with the Division of Labor." *Journal of Marriage and the Family* 50:349-61.

Berk, Sarah Fenstermaker. 1985. *The Gender Factory: The Apportionment of Work in American Households*. Plenum.

Bielby, William T., and Denise D. Bielby. 1992. "I Will Follow Him: Family Ties, Gender-Role Beliefs, and Reluctance to Relocate for a Better Job." *American Journal of Sociology* 97:1241-67.

Blair, Sampson Lee. 1993. "Employment, Family, and Perceptions of Marital Quality among Husbands and Wives." *Journal of Family Issues* 14:189-212.

Blair, Sampson Lee, and Michael P. Johnson. 1992. "Wives' Perceptions of Fairness and the Division of Household Labor." *Journal of Marriage and the Family* 54:570-81.

Blair, Sampson Lee, and Daniel T. Lichter. 1991. "Measuring the Division of Household Labor: Gender Segregation of Housework among American Couples." *Journal of Family Issues* 12:91-113.

Brayfield, April A. 1992. "Employment Resources and Housework in Canada." *Journal of Marriage and the Family* 54:19-30.

Crosby, Faye J. 1976. "A Model of Egotistical Relative Deprivation." *Psychological Review* 83:85-113.

____. 1982. *Relative Deprivation and Working Women*. Oxford University Press.

Demo, David H., and Alan C. Acock. 1993. "Family Diversity and the Division of Domestic Labor: How Much Have Things Really Changed?" *Family Relations* 42:323-31.

Hochschild, Arlie, with Anne Machung. 1989. *The Second Shift*. Viking.

Huber, Joan, and Glenna Spitze. 1983. *Sex Stratification: Children, Housework, and Jobs*. Academic Press.

John, Daphne, Beth Anne Shelton, and Kristen Luschen. 1995. "Race, Ethnicity, Gender, and Perceptions of Fairness." *Journal of Family Issues* 16:357-79.

Kanter, Rosabeth Moss. 1977. *Men and Women of the Corporation*. Basic Books.

Major, Brenda. 1987. "Gender, Justice, and the Psychology of Entitlement." Pp. 124-48 in *Review of Personality and Social Psychology*. Vol. 7, edited by Phillip Shaver and Clyde Hendrick. Sage.

Mark, Melvin M., and Robert Folger. 1984. "Responses to Relative Deprivation: A Conceptual Framework." Pp. 192-218 in *Review of Personality and Social Psychology*. Vol. 5, edited by Phillip Shaver. Sage.

Mederer, Helen J. 1993. "Division of Labor in Two-Earner Homes: Task Accomplishment versus Household Management as Critical Variables in Perceptions about Family Work." *Journal of Marriage and the Family* 55:133-45.

Perry-Jenkins, Maureen, and Karen Folk. 1994. "Class, Couples, and Conflict: Effects of the Division of Labor on Assessments of Marriage in Dual-Earner Families." *Journal of Marriage and the Family* 56:165-80.

Piña, Darlene L., and Vern L. Bengtson. 1993. "The Division of Household Labor and Wives' Happiness: Ideology, Employment, and Perceptions of Support. *Journal of Marriage and the Family* 55:901-12.

Pleck, Joseph H. 1985. *Working Wives/Working Husbands*. Sage.

Sanchez, Laura. 1994. "Gender, Labor Allocations, and the Psychology of Entitlement in the Home." *Social Forces* 73:533-53.

Shelton, Beth Anne. 1990. "The Distribution of Household Tasks: Does Wife's Employment Status Make a Difference?" *Journal of Family Issues* 11:115-35.

Sweet, James, Larry Bumpass, and Vaughn Call. 1988. *The Design and Content of the National Survey of Families and Households*. University of Wisconsin-Madison. NSFH Working Paper No. 1.

Thompson, Linda. 1991. "Family Work: Women's Sense of Fairness." *Journal of Family Issues* 12:181-96.

Thompson, Linda, and Alexis J. Walker. 1989. "Gender in Families: Women and Men in Marriage, Work, and Parenthood." *Journal of Marriage and the Family* 51:845-71.

Yogev, Sara. 1981. "Do Professional Women Have Egalitarian Marriage Relationships?" *Journal of Marriage and the Family* 43:865-71.

Ideology, Power, and Equity: Testing Competing Explanations for the Perception of Fairness in Household Labor*

ALFRED DEMARIS, *Bowling Green State University*
MONICA A. LONGMORE, *Bowling Green State University*

Abstract

Previous explanations of couples' tendency to see their household division of labor as fair primarily support the influence of three factors, net of partners' actual contributions to "female-typed" housework. The three are partners' adherence to a traditional gender role ideology, wives' limited alternatives to the marriage, and equity across key domains of the relationship. Using responses from 2,109 male and 2,096 female marital partners from the National Survey of Families and Households 1987-88, we examine the relative efficacy of these factors in accounting for variation in the perception that housework is unfair to the wife. We find most support for the effect of ideology and considerations of equity and relatively little for wives' alternatives, at least when the latter are measured by wives' general human capital. The findings suggest ways in which partners' negotiations are constrained by gender role beliefs.

Objectively, the division of household labor is clearly inequitable to women. Regardless of their employment status, wives perform, on average, more than twice as much housework as their husbands (Hartmann 1981; Presser 1994; Shelton 1990; Thompson & Walker 1989). Nevertheless, only about a third of husbands or wives perceive this situation to be unfair to wives (Lennon & Rosenfield 1994). The reasons for this apparent paradox have been investigated in a spate of recent studies (Blair & Johnson 1992; Glass & Fujimoto 1994; Lennon & Rosenfield 1994; Perry-Jenkins & Folk 1994; Piña & Bengtson 1993; Sanchez 1994; Ward 1993). Three explanations predominate for the prevalence of a sense of justice among spouses. The first is ideology: To the extent that societal norms define most household tasks as "feminine" activities, individuals

* ***This is a revised version of a paper presented at the annual meeting of the American Sociological Association, 6 August 1994, Los Angeles, CA. Data are from the National Survey of Families and Households 1987-88, which was funded by a grant (#HD21009) from the Center for Population Research of the National Institute of Child Health and Human Development.***
Address all correspondence to Alfred DeMaris, Department of Sociology, Bowling Green State University, Bowling Green, OH 43403.

will accept a division of labor in which the wife performs most of the chores. The second is power: To the extent that a woman has few alternatives to the relationship, particularly with respect to financial support, she will have little leverage with which to effect change. Her expectations for justice in housework will be correspondingly lower, thereby enhancing the apparent fairness of existing arrangements. The third is equity: To the extent that partners perceive husbands as compensating for a low input in household labor with greater inputs in other important domains of the relationship, such as paid work or child care, individuals will feel that women's greater contributions to housework are justified.

With data on more than 2,000 couples from a recent large probability sample of U.S. households, we test the relative efficacy of these three explanations in accounting for couples' perceptions of housework fairness, net of partner differences in the actual performance of housework. Both married and unmarried cohabiting couples are employed in the analyses; the terms *husband* and *wife* will be used throughout to refer generically to unmarried as well as married intimate partners. Unlike several other analyses, we examine both husbands' and wives' perceptions of housework justice, rather than just those of the wife. Also, in contrast to most previous studies on this topic, we examine the impact of partners' relative contributions to child care on perceptions of fairness in housework. Child care represents one of several domains of the relationship other than housework per se that, under an equity model, may affect perceptions of fairness. We begin by reviewing theoretical and empirical literature pertaining to the perception of housework fairness.

Theoretical Background

HOUSEWORK AND GENDER

Household work is a gendered activity, particularly for those in marital or quasi-marital relationships (South & Spitze 1994). Not only is the bulk of housework performed by women, but these tasks are also allocated disproportionately according to gender. Women are chiefly responsible for tasks such as cooking, washing dishes, doing laundry, and cleaning house, and men tend to be responsible for tasks such as home repair, automobile maintenance, and yard work. Other jobs such as grocery shopping or paying bills appear to be more evenly divided between the sexes (Bergen 1990; Nyquist et al. 1985; Ross 1987). Following Schooler et al. (1984), we refer to "female typed" tasks as "routine" and all others as "sporadic," to highlight the differences in time commitment that characterize these tasks: Routine work must ordinarily be done daily, whereas other tasks allow greater flexibility in scheduling. Women additionally have the primary responsibility for the planning and supervision of such tasks. This "executive" responsibility has led Schooler and his associates to observe that for women, unlike men, housework "is an imperative with psychological effects similar to those of the structural imperatives of paid employment" (122).

A "just" division of labor in the home would therefore be one in which men shared equally both in the performance of *all* household chores and in the

responsibility for seeing to it that housework is done (Perry-Jenkins & Folk 1994). Here justice is defined in the narrow sense as the equality of partners' contributions to each type of work required to sustain the household. (Below we argue that equity — a form of justice that subsumes equality as a special case — may be as influential, if not more so, in partners' perceptions of "justice.") Men's willingness to contribute equally to this workload is held to be important to women for several reasons. Obviously, if men contribute more to the performance of some of the relentless daily tasks of the household, women's leisure time would be comparable to men's. Perhaps of greater significance, however, is the symbolic value of such sharing. Housework has been characterized as "menial work performed by a lower status person for someone of higher status" (Ross, Mirowsky & Huber 1983:821). When husbands share "women's" tasks, on the other hand, neither spouse should feel that he or she is "serving" the other, implying that theirs is a relationship between equals. For these reasons, we expect that *the greater men's relative contribution to the routine tasks of the household, the less likely either spouse is to see housework as unfair to the wife.*

Several studies have indeed found that housework was perceived as unfair to the wife the greater the proportion of routine, or "female-typed," housework done by her (Blair & Johnson 1992; Lennon & Rosenfield 1994; Perry-Jenkins & Folk 1994). Moreover, when husbands' relative contributions to both routine and sporadic housework have been included in analyses, only the former have shown a significant effect on wives' sense of fairness (Blair & Johnson 1992; Lennon & Rosenfield 1994). For a given division of labor in the home, however, other factors are also likely to influence perceptions of fairness.

IDEOLOGY

Through internalizing social norms of gendered behavior, individuals form expectations about their proper roles in marriage. Traditional norms encourage women to be more nurturant than men and to care more about cleanliness and hygiene. Traditional socialization introduces girls — more than boys — to housework at an early age, thereby promoting anticipatory socialization for the later role of "housekeeper." More modern notions suggest, of course, that in the spirit of gender equality, either spouse is equally able to perform both routine and sporadic housework.

We therefore expect that *the more spouses subscribe to an egalitarian sex-role ideology, the more they will tend to see the existing arrangement — in which women do more of the routine tasks — as unfair to wives.* That is, holding constant men's relative contribution to routine housework, we expect that those with more egalitarian beliefs will have higher expectations for men's contributions. Because men's performance typically lags behind expectations in the area of housework, those with greater expectations should be more likely to see existing arrangements as unfair to the wife. Research to date, however, has produced mixed results for the impact of ideology. Some studies have found that wives subscribing to greater gender role egalitarianism see the household division of labor as more unfair to them (Blair & Johnson 1992); while others have found that ideology has no effect (Lennon & Rosenfield 1994). Piña and Bengtson's (1993) data show a more traditional gender ideology to be predictive of a

greater likelihood of being satisfied with the "help and support received" from husbands, but only for wives who are not employed full-time (Piña & Bengtson 1993).

POWER

Lennon and Rosenfield (1994) have recently articulated a social exchange perspective that suggests that key determinants of perceived fairness in household labor are women's alternatives to remaining in the relationship. The fewer a wife's alternatives to the relationship, the more dependent she is on it, and therefore the less power she has with respect to her spouse. Women in these straits have limited expectations for justice. Because they have little leverage with which to induce more cooperation in housework from their husbands (England & Farkas 1986), they instead redefine existing arrangements as "fair" to maintain cognitive consistency between behavior and belief. Moreover, they tend to make social comparisons with other wives who are similarly powerless. Hence, they may see their division of labor as satisfactory in spite of considerable objective inequities in housework, simply because they do not feel entitled to more. Some support for these dynamics has been found in the study by Lennon and Rosenfield (1994). They show that controlling for partners' contributions to housework and sex-role traditionalism, women who perceive fewer options to the current relationship and women whose earnings are below the poverty threshold tend to see the division of labor as more fair to them. This suggests that *the more alternatives women have to the current relationship, the more they will define the division of labor as unfair to them.*

One problem with this hypothesis is that, at any given point in a marriage, the existing division of labor is the end product of a process of negotiation that is largely transparent, particularly in cross-sectional data. As Thompson (1991) suggests, women may not feel that the existing division of labor is unjust if they "accept as appropriate the participatory process by which the outcomes were created" (193). Where women's alternatives should really make a difference is in the negotiations over how much each partner will contribute to household tasks. One would expect that women with more alternatives will have correspondingly greater power to negotiate a balance of housework that is more favorable to them (Blumberg & Coleman 1989). And, regardless of who is favored objectively in the division of labor, wives with more power should be more likely to participate in decisions about the extent of their housework contribution. With this in mind, one could entertain the opposite hypothesis that *existing housework arrangements will be perceived as fairer by the more powerful than by the less powerful wives.* The former are more likely both to have participated in the decision-making process that led to the current outcome and to have effected outcomes consistent with their preferences. Indeed, to the extent that couples experience greater mutual influence in negotiations, they tend to see the resulting behavioral arrangements as more equitable (Scanzoni & Godwin 1990). We will allow the data to determine which of these divergent hypotheses is the more plausible.

EQUITY

Definition

Equity is said to be achieved in an exchange whenever each partner's outcomes are proportional to their inputs (Adams 1965; Sprecher 1986; Walster, Walster & Berscheid 1978). Outcomes are the rewards one receives from an exchange. Inputs are one's own contributions to the exchange (Walster, Walster & Berscheid 1978). In an intimate relationship, a husband's input is usually the wife's outcome, while the wife's input is, correspondingly, the husband's outcome (Cook & Hegtvedt 1983). Formally, equity is achieved whenever (Adams 1965; Jasso 1978)

$$\frac{\mathit{Outcome(Husband)}}{\mathit{Input(Husband)}} = \frac{\mathit{Outcome(Wife)}}{\mathit{Input(Wife)}} \tag{1}$$

Equity across Marital Domains

It is rarely the case that partners either make absolutely no contribution to a particular domain of their relationship or that they make exactly equal contributions to each domain. Instead, partners usually make some contribution to every aspect of the relationship, but one partner's input in a given area usually outweighs the other's. Hence, outcome/input ratios in a given area, such as time spent in routine tasks, are usually unequal. Nevertheless equity can be achieved if an unequal ratio in one domain — such as, again, routine housework — is counterbalanced by an unequal ratio in another domain — such as paid work — in the opposite direction. Therefore, both partners may feel it equitable if the wife, say, does more routine housework but the husband spends more time in paid labor.

Factors Contributing to Equity

Intimate relationships involve the mutual transfer of a wide range of both tangible and intangible rewards. Assessing whether such transfers are balanced is made difficult by the fact that individuals often place different values on the same commodity. At the same time, it is possible to measure outcome/input ratios for a given commodity whenever such commodities are (1) more tangible and (2) measured in the same metric for each partner. With respect to the tangible contributions that are relevant to the total work of a household, three are paramount: (1) contributions to housework, (2) contributions to child care, and (3) contributions to standard of living. Housework entails both routine and sporadic tasks, as outlined above. Child care entails both the additional housework involved in feeding and clothing children and activities connected with child socialization.

Several types of contributions by partners are relevant to the standard of living of the household. These are income, the number of hours in paid labor that is required to obtain that income, and education, which indexes both the status of the household and the cultural milieu in which children are raised. Taking on a second job also indexes a commitment to provide support for the household, beyond what is contributed through one's primary employment. Health status is an additional factor that would be expected to be related to

work performance, whether housework or paid labor, and therefore provides an important indirect contribution to a household's standard of living. When a spouse is in poor health, his or her energy is usually reduced, particularly for the performance of physical labor. Should either partner be impaired, the other partner will be obliged to "pick up the slack" in the total workload of the household. Hence, the relative health status of the partners is important to relationship equity, because the partner in better health may feel that he or she is forced into doing more than his or her "share."

In sum, equity theory leads us to expect that, net of the husband's relative contribution to routine housework, *(1) the greater the husband's contribution to sporadic housework, child care, income, hours in paid labor, education, and health, relative to his wife's and (2) when the husband has a second job but the wife does not, both partners are less likely to see the division of labor as unfair to the wife.*

A few studies have examined the perception of housework fairness as a function of spouses' relative educational or income levels, or wives' hours in paid labor (Benin & Agostinelli 1988; Glass & Fujimoto 1994; Lennon & Rosenfield 1994; Sanchez 1994; Ward 1993). Little support for equity theory has been found thus far. Nevertheless, equity theory leads us to expect some congruence between objective indicators of partners' relative contributions to the domains enumerated above and subjective perceptions of housework fairness. Below we consider why previous research has failed to find such links.

Limitations of Previous Research

There are several limitations in studies of housework fairness to date that either constrain their generalizability or prevent us from establishing which competing theory receives the most support. For example, several studies examine only wives' perceptions of housework fairness (Blair & Johnson 1992; Glass & Fujimoto 1994; Lennon & Rosenfield 1994; Piña & Bengtson 1993), thereby preventing comparisons across gender. Other studies restrict their population to dual-earner couples (Benin & Agostinelli 1988; Lennon & Rosenfield 1994; Perry-Jenkins & Folk 1994), again preventing generalizability to couples with stay-at-home wives. Perhaps more important, with the exception of Lennon and Rosenfield's (1994) study, researchers have typically not included factors representing ideology, power, and equity simultaneously in their models. Consequently, it is not possible to examine the independent contribution of each factor in accounting for perceptions of fairness. Moreover, when equity effects are included, relatively few domains of the relationship have been represented. Of the omitted areas, child care looms as especially important, as husbands may partially balance a low contribution to housework by doing a greater share of child care (Spitze 1988).

Some studies have used as independent variables factors that may themselves be the *effects* of perceived fairness in housework. For example, Lennon and Rosenfield (1994) measure a woman's perceived alternatives to the relationship, or "options," as her perception that, upon separation from the spouse, her standard of living, social life, and career opportunities would improve. It is entirely possible, however, that those who see their division of

labor as unfair will consequently come to see themselves as "better off" without the current spouse. If the causal order of the variables is reversed, or even if there are reciprocal effects, the estimate of the effect of the "independent" variable will be biased. The questionable nature of this measure is even more evident when "options" are used as a predictor of depressive symptomatology. Women who have options to the current marriage are believed to be empowered as a result (Lennon & Rosenfield 1994). Given relationships that are equally happy, one would expect that women with more power will be less depressed. However, Lennon and Rosenfield (1994) found that, controlling for the perceived fairness of housework and marital happiness, women with more options were significantly *more* depressed. These effects call into question the validity of such a measure as an index of women's alternatives to the current marriage.

The present study attempts to redress these problems by (1) examining the perceptions of both husbands and wives, (2) including all couples in which at least one partner is in the work force full-time, (3) including measures of partners' relative contributions to child care, (4) employing measures of ideology, power, and equity simultaneously in models of housework fairness, and (5) employing independent variables with unambiguous causal relationships to perceived fairness.

Methods

THE DATA

The data for this study were drawn from the National Survey of Families and Households 1987-88 (NSFH), a multistage probability sample of 13,008 adults in the noninstitutionalized U.S. population. One individual per household was randomly selected as the primary respondent for face-to-face interviewing. A shorter, self-enumerated questionnaire was given to spouses and cohabiting partners when present. Several subpopulations were oversampled: unmarried cohabitors, recently married couples, single-parent families, families with stepchildren, and households whose heads were African American, Mexican American, or Puerto Rican (Sweet, Bumpass & Call 1988).

Because questions about partners' contributions to the socialization of children were asked only of respondents with children between 5 and 18 in the household, our sample is necessarily limited to that group. We further restricted our sample to households in which the primary respondent was married and living with the spouse, or cohabiting unmarried, and the partner's questionnaire was returned completed. Moreover, because our theoretical arguments are couched in terms of gendered behavior, we excluded same-sex couples. Couples in which either partner was in the armed forces and households in which neither partner was employed full-time were also eliminated. The dynamics of household justice are likely to be different among those who are either part of the military subculture, or among whom neither partner is in the work force full-time. Our remaining sample of 2,153 couples is therefore reasonably representative of nonmilitary heterosexual couples with children among whom at least one partner is in the work force full-time. In that analyses were carried

out separately for husbands and wives, we excluded individual partners with missing data on the dependent variable. Our final sample therefore consists of 2,109 husbands and 2,096 wives.

MEASURES

The Dependent Variable

The perception of fairness in the division of household labor, the dependent variable, was tapped with a single item from the NSFH that asked "How do you feel about the fairness of household chores in your relationship?" Possible response categories were "very unfair to me," "somewhat unfair to me," "fair to both," "somewhat unfair to her/him," and "very unfair to her/him." These responses were coded from 1 to 5, respectively, and constitute a scale with high scores indicating overbenefit. However, for reasons detailed below, a dichotomized version of this variable was also created, with the perception that chores were very or somewhat unfair to the wife contrasted with all other responses.

The Independent Variables

The predictors can essentially be grouped into five categories: partners' relative contributions to routine housework, ideology factors, wives' alternatives, equity factors (including contributions to sporadic housework, as explained below), and controls. Husbands' and wives' contributions to household labor were measured by the number of hours per week that each partner normally spends doing the following nine tasks: (1) cleaning house; (2) preparing meals; (3) washing dishes and cleaning up after meals; (4) washing, ironing, and mending; (5) automobile maintenance and repair; (6) paying bills and keeping financial records; (7) shopping for groceries and other household goods; (8) outdoor and other household maintenance tasks; and (9) driving other household members to work, school, or other activities. Both the main respondent and his or her partner were asked to provide this information; therefore there are two perspectives available on the amount of housework done by each partner, for a total of four sets of responses. Exploratory factor analyses with varimax rotation revealed that, regardless of whose reports were scrutinized, and regardless of whether it was the husband's or the wife's hours in question, two dimensions emerged, with items 1-4 loading primarily on one factor, and items 5-9 loading primarily on the other.

From these considerations we summed the first four items to form a *routine housework* scale and the last five items to form a *sporadic housework* scale.[1] The reliability of these scales varied, depending upon which partner provided the information, but ranged from a low of .54 for females' reports of males' hours in sporadic tasks, to a high of .82 for males' reports of females' hours in routine tasks. Average reliability over the eight scales (four based on each partner's reports) was .71. Husbands' reports of both partners' housework hours in all nine tasks were employed to create the scales for the analysis of husbands' perceptions of housework fairness. Similarly, wives' reports were used to create scales for the analysis of wives' perceptions of fairness. In that it is the respondents' impression of how much time each partner spends doing housework that is likely to influence their sense of justice in the division of

labor, it was important to employ their own perceptions of housework time in each case. Missing values on any item were replaced by the mean for that item, prior to summing.[2]

Two ideology variables were created for the analyses. The first was a *sex-role egalitarianism* scale consisting of six items from the NSFH. The first four asked for the extent of agreement with these statements: (a) It is much better for everyone if the man earns the main living and the woman takes care of the home and family, (b) Preschool children are likely to suffer if their mother is employed, (c) Parents should encourage just as much independence in their daughters as in their sons, and (d) In a successful marriage, the partners must have freedom to do what they want individually. Responses ranged from strongly agree (1) to strongly disagree (5). Items (c) and (d) were reverse-coded so that high scores on these items indicated the most egalitarian beliefs. The other two items asked respondents the extent of their approval of (e) Mothers who work full-time when their youngest child is under age 5 and (f) Mothers who work part-time when their youngest child is under age 5. Responses on these items ranged from strongly approve (1) to strongly disapprove (7), and were also reverse-coded so that the high score represented the most egalitarian response. Because of different response formats, all items were standardized before being summed to create the scale. Scale reliabilities were .66 for husbands and .67 for wives. *Endorsement of housework sharing* was the second ideology factor and was included to tap egalitarianism specifically in the sharing of housework. The single item consisted of responses to the statement, "If a husband and wife both work full-time they should share household tasks equally." Responses ranged from strongly agree (5) to strongly disagree (1). Both husband's and wife's endorsements of sharing were tapped. The foregoing seven items have been used in various combinations to tap gender role ideology in other studies (Blair & Johnson 1992; Lennon & Rosenfield 1994; Presser 1994; Szinovacz & Harpster 1993).

In their direction of influence, these items are considered to be antecedent to perceptions of the fairness of housework, although it is conceivable that seeing housework as unfair in one's relationship might lead one to adopt a more egalitarian stance on gender roles. However, we think it more reasonable that gender ideology precedes evaluations of justice in the marriage. Sex-role preferences are most likely crystallized during late adolescence and young adulthood (Morgan & Scanzoni 1987), as individuals engage in anticipatory socialization for marriage. Moreover, the recognition of injustice presupposes a preexisting standard against which the current arrangement is compared. That standard is a set of beliefs about the proper division of marital roles between partners.

Wives' alternatives to the marriage comprise both subjective and objective dimensions. Because of the previously mentioned problems with investigating the primary subjective measure of dependency in this data set — perceived life change upon separation from the partner — we have elected instead to measure wives' objective alternatives. Scholars have argued that wives have alternatives or "options" to the current marriage to the extent that they are able to financially support themselves outside it (Becker 1991; England & Farkas 1986; Lennon & Rosenfield 1994). To measure this, we include four factors representing the

wife's store of general human capital (Becker 1991; England & Farkas 1986): *income, educational level, hours employed in the labor force,* and *self-assessed health.* We expect that all else being equal, wives with greater income, higher levels of education, a greater commitment to full-time employment — indexed by the number of hours currently employed — and better health will be more able to survive economically outside the marriage than will others, and therefore they have more alternatives to it.

The measure of income was the respondent's total income, excluding income from interests, dividends, and other investments. This variable therefore primarily reflects individuals' earnings from employment. Education was measured in years of formal schooling. Hours currently employed were coded from several questions in the NSFH on employment. Respondents who were currently working for pay were asked how many hours they worked the last week. If that was the number of hours they usually worked, it was used as their hours of employment. If not, hours employed were coded as the typical number of hours worked per week. Those who were not currently employed or who had missing responses for hours of employment were assigned a value of zero on this variable. Health status was based on respondents' reports of the state of their health compared to other people of the same age. Their responses, ranging from "very poor" (1) to "excellent" (5), were treated as interval-level independent variables. Self-reported assessments of health demonstrate good criterion validity, predicting survival over a 12-year period for both sexes (Idler & Angel 1990). Missing values for education, income, and health were replaced by means on those variables.

Husbands' and wives' *relative statuses* on seven factors were included in order to assess the extent to which perceptions of fairness are affected by equity considerations. The first was contributions to *sporadic housework.* This was tapped by the sporadic housework scales described above. One might argue that this is simply part of the total housework load and therefore should be considered on a par with routine housework. However, that would be contrary to a large body of theory and research that suggests that it is only the time that men spend doing "female-typed" tasks that helps to balance out the gender gap in the division of labor (Benin & Agostinelli 1988; Blair & Johnson 1992; Perry-Jenkins & Folk 1994). Our data in fact show that, although men's hours in routine and sporadic housework are about the same (7.3 and 11.4 hours, respectively, according to wives' reports), women spend about three times as many hours doing routine tasks as they spend on sporadic tasks (33 and 11 hours, respectively). Therefore men's time in sporadic housework is not equivalent to women's time in routine housework. Rather, it "counts" toward the fairness of housework only by being considered an "equitable" contribution.

The second through the fifth factors were the human capital factors just enumerated: *income, education, hours in paid employment,* and *self-assessed health.* The sixth factor was *having a second job.* Dummy variables were created to represent whether either the wife or the husband worked at a second job. We consider the wife's having a second job as relevant to equity in the relationship, rather than tapping her alternatives. Although hours of employment index commitment to working outside the home, net of this, having a second job

typically indicates a low-paying primary job rather than a commitment to employment, per se.

The seventh factor was contributions to child care. For households with children, two aspects of child care are paramount. The first consists of the extra housework that must be performed to feed, clothe, and otherwise provide an adequate living environment for children. These chores are already subsumed under the total time spent in routine and sporadic housework. The other consists of activities pertinent to child socialization, such as playing with children, teaching them table manners, helping them with homework, engaging with them in recreational pursuits, and so forth. Child socialization activities were measured in the NSFH by six items asked of all respondents with one or more biological, step-, adopted, or foster children 5-18 years old in the household. Each partner was asked (a) how many days "last week" he/she ate breakfast with at least one of the children, (b) how many days "last week" he/she ate dinner with at least one of the children, (c) how often he/she engaged in leisure activities with the children away from home, (d) how often he/she spent time at home with the children working on a project or playing together, (e) how often he/she had private talks with the children, and (f) how often he/she helped the children with reading or homework. Items (c) through (f) were coded from 1 (never or rarely) to 6 (almost every day).

Factor analyses of husbands' and wives' responses — again employing varimax rotation — suggested a two-factor model for these items, with one factor represented by items (a) and (b) and the second factor represented by items (c) through (f). Therefore we created summed scales for both husbands and wives representing these factors, called respectively *mealtime with children* ($\alpha = .40$ for husbands, .47 for wives) and *child socialization activities* ($\alpha = .75$ for husbands, .74 for wives). Missing values for any item were replaced in each case by the item mean. Some construct validation for these items was provided by examining whether the relative contribution to child care affected respondents' perceptions of the fairness of child care in their relationship. As one would expect, significant relationships were observed between the fairness of child care — coded as unfair to the husband, fair to both, or unfair to the wife — and partners' relative contributions to both mealtime and child socialization. Husbands and wives were more likely to see child care as fair to both partners if their contributions were equal or if the husband did more, than if the wife contributed the greater share of either mealtime or child socialization activities. (Below we discuss in greater detail how partners' relative positions on these seven items were entered into the model.)

Several variables were included in the predictor set as controls because of their potential for confounding relationships between the explanatory variables of primary interest and the dependent variable. For example, older couples, those with longer marriages, and those who are married rather than cohabiting unmarried are likely to be both more conventional in gendered behavior and less likely to see housework as unfair to the wife. Failure to control for them may result in an inflated estimate of the impact of ideology on housework fairness. For these reasons we controlled for marital duration and husband's age in years, cohabiting status (married is the contrast category), and number of children 5-18 years old. To capture race differences, we entered three dummy

variables representing the ethnicity of the couple. The three categories were (1) both partners were black, (2) both partners were Hispanic, and (3) the two partners were both from other ethnic groups — primarily Asian or American Indian — or were each from a different ethnic group. Both partners being white was the contrast category. We also included three dummy variables indicating the relationship of children in the household to each partner. We therefore distinguished stepfather families, stepmother families, and families characterized by other parent-child relationships (primarily double stepfamilies and those with adopted or foster children) from families in which all children were the biological children of both partners (the contrast category). Of the controls, only marital duration was missing for some of the cases, in which case the mean value was substituted.

STATISTICAL MODEL

The perception of housework fairness, our dependent variable, has frequently been treated as an interval scale in other analyses (Glass & Fujimoto 1994; Lennon & Rosenfield 1994). We feel that this may not be the best strategy, for two reasons. First, virtually all scholarly writing on housework suggests that, if anything, the division of household labor is unfair to women. Hence our interest really centers on the degree to which partners see housework as unfair to the wife, rather than the husband. Indeed, an examination of the sample distribution on the dependent variable reveals that fewer than 5% of either husbands or wives consider housework to be unfair to the husband (the exact figures for the unweighted distributions are 3.6% of husbands and 4.8% of wives). Perhaps more important, however, is that the distributions are strongly skewed in the direction of fairness to both partners or unfairness to the wife: more than 95% of responses on this five-level variable fall in the range of 3-5 for husbands and 1-3 for wives. When ordinal variables have few categories, and their distributions are skewed, applying ordinary least squares (OLS) regression may be somewhat problematic. The constant variance assumption required for least-squares estimators to be "best" is violated by categorical data (Agresti 1989), particularly when the distribution across categories is very uneven, as in this case. A consequence is that OLS estimators are inefficient, making for less sensitive tests of individual parameters.

Given these considerations, we employed logistic regression to model the log odds that housework is considered unfair to the wife — as opposed to fair to both or unfair to the husband — as a linear function of the predictor set. At the same time, we also ran the model using OLS with the full five-level dependent variable for comparison purposes. OLS was also employed to diagnose the X-matrix for potential collinearity problems (in SAS or SPSS these diagnostics are available only in the linear regression procedure) and to assess whether our missing value imputation procedure resulted in bias. Comparisons of the OLS and logistic regression results produced no substantive differences. All the significant effects in either type of analysis were in the same direction and supported the same conclusions. However, five more effects were significant across the two (husbands and wives) logistic regressions, that were not significant in the OLS analyses. This tended to confirm the lower sensitivity of

OLS in this instance. For this reason we present the logistic regression results of our multivariate analyses below.

To explore whether collinearity was a problem in our data, we examined the variance inflation factors (VIFs) for each coefficient in the OLS models for both husbands and wives. VIF values above 10 generally indicate collinearity problems (Myers 1986). In our estimated equations, all the VIFs were well under 6. Evidently, collinearity was not severe enough to pose any major problems in our analyses. As part of our diagnostic tests, we also created dummy variables indicating whether any missing values had been imputed for any of the variables for either husbands or wives. Adding these to the models produced virtually no detectable changes in the coefficient values. Nevertheless, it was possible that the impacts of the predictors might be different for those with values imputed, compared to those with complete data on all variables. We checked this by adding a set of terms representing the interactions between whether missing values were imputed values and all other predictors, to the base model. For both husbands and wives, the set of interaction terms was very insignificant ($p > .4$). It appears that imputed missing values did not substantially affect our results. Our multivariate models are therefore presented without controls for imputation. All inferential analyses (including the bivariate results discussed below) have been weighted to correct for oversampling. Weights have additionally been modified to sum to the unweighted number of cases.

MODELING EQUITY ACROSS MARITAL DOMAINS

Authors of previous studies have employed a number of strategies for tapping partners' relative contributions to the marriage, including ratios (Ward 1993) and logged ratios (Presser 1994) of partners' resources and a given partner's proportionate contributions in different areas (Lennon & Rosenfield 1994). Each strategy presupposes a specific conceptualization of how equation 1 above is to be modified to reflect balance across the multiple domains of marriage. Of the three techniques just mentioned, we prefer an additive model of logged ratios of partner's resources. In contrast to proportionate contributions, they imply a very straightforward extension of equation 1. Formally, if O_{jh} and I_{jh} are the husband's outcome and input, respectively, in the jth domain of the relationship, and O_{jw} and I_{jw} are the corresponding outcome and input of the wife in the jth domain, then we can envision equity obtaining across the domains of a relationship if

$$\left(\frac{O_{1h}}{I_{1h}}\right)\left(\frac{O_{2h}}{I_{2h}}\right)\cdots\left(\frac{O_{Jh}}{I_{Jh}}\right)=\left(\frac{O_{1w}}{I_{1w}}\right)\left(\frac{O_{2w}}{I_{2w}}\right)\cdots\left(\frac{O_{Jw}}{I_{Jw}}\right) \tag{2}$$

That is, equity is achieved whenever the products of the outcome/input ratios across the $j = 1, 2, \ldots, J$ domains of the relationship are the same for husbands and wives. Taking natural logs of both sides, equation 2 implies

$$\sum_{j=1}^{J} \log\left(\frac{O_{jh}}{I_{jk}}\right)=\sum_{j=1}^{J} \log\left(\frac{O_{jw}}{I_{jw}}\right) \tag{3}$$

This formulation suggests that equity is obtained to the extent that the sums of logged husband-wife ratios of contributions to each key domain of the relation-

ship approach equality across partners. In terms of the fairness of housework, this formulation also suggests that the greater the right side of equation 3, and the lower the left side, the less likely either partner will be to perceive housework as unfair to the wife.

An example will help to clarify how equations 2 and 3 capture equity/inequity across domains. We assume that husbands' inputs are wives' outcomes, and vice versa. Suppose that the husband spends 50 hours per week in paid labor, does 10 hours of housework, has a college degree, and makes $50,000 per year. His wife, on the other hand, spends 40 hours in paid labor, does 30 hours of housework, has 3 years of college, and makes $30,000 per year. The left side of equation 2 is: (40/50) (30/10) (15/16) (30,000/50,000) = 1.35, while the right side is: (50/40) (10/30) (16/15) (50,000/30,000) = .741. The exchange is clearly imbalanced in favor of the husband. If, on the other hand, his income were $67,500, both sides of equation 2 would equal one, and the exchange would be equitable.

Miles's (1987) review of the equity formulas proposed by various theorists enumerates a number of criteria that an equity formula should satisfy to be valid. Equation 2 satisfies those criteria provided that both outcomes and inputs in each ratio are nonnegative and inputs are bounded away from zero. Both conditions are satisfied in this study. None of the resources used to create ratios can logically take negative values. For resources that could take on a value of zero (e.g., hours in paid labor), one-half was added to the numerator and denominator of the ratio prior to taking the log to prevent the occurrence of undefined logs or undefined ratios. Equation 2 also effectively allows combining outcomes or inputs measured in different units, a problem that plagues other formulations (Miles 1987). Because the same units of measurement cancel in the numerator and denominator of ratios, they constitute a unitless metric. Although Farkas and Anderson's (1979) equity integration formula also proposes to take account of multiple inputs, it is quite cumbersome compared to equation 2. Moreover, it is not clear how their formula is to be extended to the case of multiple outcomes and multiple inputs simultaneously (Miles 1987). In our formulation, on the other hand, that extension is straightforward.

We employ logged husband-wife ratios of sporadic housework hours, mealtime with children, child socialization activities, income, hours in paid labor, and health status to measure equity in these domains. Because relative contributions to routine housework are also part of the overall equity equation, they were also tapped using the log of the husband-wife ratio. An examination of the logged ratio of partners' educational levels revealed an extremely leptokurtic distribution with little variance and only a few observations reflecting large educational differences. We therefore assessed partners' relative education with dummy variables indicating whether either the husband or the wife had more education (equal education was the contrast). As previously noted, dummy variables were also used to indicate whether the husband or the wife had a second job. Finally, for each of the logged ratios, as well as relative educational standing, we also included in the model the wife's corresponding contribution in raw units. This allows us to separate the effects on housework fairness due to wives' absolute contributions across domains from those due to

partners' *relative* contributions. Descriptive statistics for all variables in the analysis are shown in the Appendix.

Results

BIVARIATE ASSOCIATIONS

Before proceeding to the evaluation of the multivariate models, we briefly consider some bivariate findings of interest. In Table 1, we present a comparison of the proportions of husbands and wives who consider housework to be, alternately, unfair to the wife, fair to both partners, and unfair to the husband. For this analysis we look only at the 2,054 couples in which both partners responded to the question on fairness. Because these are dependent proportions, we test for gender differences using McNemar's chi-squared test (Agresti 1990). It is evident that wives are more likely than husbands to see housework as unfair to women. About 36% of wives as opposed to 30% of husbands give this response. Correspondingly, husbands are more disposed to see housework as fair to both (66.3%) than are wives (59.4%). Husbands and wives appear to be in harmony on whether housework is unfair to husbands: only 3.7% of husbands and 4.6% of wives give this response, an insignificant difference.

Tables 2 and 3 show the relationship between husband's relative contributions to each type of housework and the perceived fairness of housework. The association is shown from two different perspectives. Table 2 is similar to Table 4 in Lennon and Rosenfield (1994).[3] As in their table, ours examines the conditional distribution of the husband's relative contribution to housework, given husbands' and wives' reports of housework fairness. Our measure of relative contribution is the average ratio of husband's and wife's hours of housework. In this case, perceptions of fairness are — incorrectly — being treated as the "independent" variable. (Customarily, one examines the conditional distribution on the dependent variable, given levels of the independent variable.)

We see from the table that, among both husbands and wives who see housework as fair to both partners, husbands perform, on average, between 11 and 22% more of the sporadic chores than wives do (the ratios are 1.108 using wives' reports, and 1.217 using husbands' reports). Among those who see housework as unfair to the husband, husbands do between 13 and 34% more of the sporadic chores. Among those who see housework as unfair to the wife, husbands only do 85 to 95% as much sporadic housework as their wives. Lennon and Rosenfield (1994) correctly conclude from this that both women and men appear to believe that men should do more than half of the male-typed tasks.

Turning to the routine tasks, we note the striking imbalance between men's and women's actual task performance, regardless of how fair it is considered to be. For example, both husbands and wives who see housework as unfair to husbands estimate that men only do between 36 and 39% as much as women. Those who say it is fair to both estimate men's performance to be only 19% of women's. And those who see housework as unfair to the wife see men as doing only 13 to 14% of the housework that women do. Lennon and Rosenfield's

TABLE 1: Tests for Gender Differences in Perceptions of Housework Fairness[a]

	Respondent Husband	Respondent Wife	McNemar's χ^2
Perception that housework is			
Fair to both	.663	.594	27.102*
Unfair to the husband	.037	.046	2.729
Unfair to the wife	.300	.359	21.492*
(N = 2,054 couples)			

* $p < .001$

conclusion from these findings is that they reflect "a general *acceptance of an unequal division of housework,* with women spending many more hours on the routine, time-consuming, traditionally female tasks" (Lennon & Rosenfield 1994:520-21, emphasis added).

To show why this conclusion is ill-advised, we present Table 3, which correctly shows the conditional distribution on housework fairness, given husbands' and wives' relative performance of housework. For this analysis, the ratios of husbands' to wives' housework hours have been collapsed into three categories: the husband does more housework, both partners do the same amount, and the wife does more housework. Because so few partners are exactly equal in housework hours (e.g., only 14 are exactly equal in routine housework, according to wives), we have considered contributions to be equal if the husband's contribution is within 10% of the wife's (i.e., the ratios are between .9 and 1.1). In our sample, in which wives perform, on average, about 33 hours of routine housework per week, this amounts to an average weekly gender difference of 3.3 hours of routine housework either way, or fewer than 30 minutes per day. Other researchers have considered the division of labor to be "equal" as long as the gender difference is less than one hour per day (e.g., Piña & Bengtson 1993).

Table 3 shows that the likelihood of saying that routine housework is fair to both partners is *highest* among couples in which both partners do approximately the *same* amount of housework and lowest when the *wife* does more. The findings with respect to sporadic housework confirm the conclusions from Table 2; that is, partners are most likely to see housework as fair when the husband does more of these tasks. Why are the conclusions about partners' expectations for fairness so different for routine housework, when Table 3 rather than Table 2 is examined? In that women do more of the routine tasks in more than 93% of the couples in the sample, examining the conditional distribution of partners' relative contributions to housework across categories of fairness accurately reflects the gender gap in housework *performance,* regardless of the acceptance (i.e., perceived fairness) of the arrangement. However, examining the perception of fairness, conditional on housework performance (as in Table 3), it

TABLE 2: Mean Husband/Wife Housework Ratios by Perceptions of Housework Fairness, for Husbands and Wives

	Mean Ratio for Those Who Say Housework Is:			
	Unfair to Husband	Fair to Both	Unfair to Wife	F[a]
Husbands				
Routine tasks	.393	.190	.128	36.740*
Sporadic tasks	1.344	1.217	.946	25.280*
(N = 2,109)				
Wives				
Routine tasks	.358	.187	.143	21.180*
Sporadic tasks	1.126	1.108	.845	32.610*
(N = 2,096)				

[a] Test for equality of the means of the logged ratios, based on one-way ANOVA

* $p < .001$

is clear that both husbands and wives are most likely to *accept* the situation as fair when housework is equally apportioned.

MULTIVARIATE FINDINGS

Tables 4 and 5 present the results of logistic regressions of the log odds of seeing housework as unfair to wives on the predictor set, separately for husbands and wives. To conserve space, we show only the significant predictors and exclude the intercept. (Full tables are available from the authors upon request. See the Appendix for a listing of the complete predictor set used in the multivariate analyses.) Shown at the bottom of the tables are chi-squared tests for the set of predictors in the model; both are highly significant. We also show McKelvey and Zavoina's (1975) R^2 analog for logistic regression, denoted by $\hat{R}^2$. This measure is an estimate of the proportion of variance that would be accounted for by the model in unfairness to the wife, if the latter were measured as a continuous variable. These values are .21 for husbands and .16 for wives.

TABLE 3: Perceptions of Housework Fairness by Relative Contributions to Housework, for Husbands and Wives

	Relative Contribution to Housework			
	Husband Does More	Both Do the Same	Wife Does More	χ^2
Perceptions of fairness				
Husbands				
Routine tasks				
Unfair to husband	.187	.145	.028	
Fair to both	.729	.769	.659	
Unfair to wife	.084	.087	.314	90.621*
Sporadic tasks				
Unfair to husband	.042	.023	.032	
Fair to both	.715	.613	.597	
Unfair to wife	.243	.364	.372	40.867*
(N = 2,109)				
Wives				
Routine tasks				
Unfair to husband	.170	.094	.042	
Fair to both	.716	.740	.588	
Unfair to wife	.115	.166	.371	40.700*
Sporadic tasks				
Unfair to husband	.055	.043	.037	
Fair to both	.658	.606	.508	
Unfair to wife	.287	.351	.455	52.128*
(N = 2,096)				

* $p < .001$

Results for Husbands

Of the controls, husbands' age, the race of the couple, and the nature of the relationship between the children and each partner all have significant effects on husbands' perceptions of whether housework is unfair to wives. The older the husband, the less likely he is to consider housework to be unfair to his partner. In that gender-role attitudes are controlled in the analysis, this effect is over and above the association of increasing age with gender-role traditionalism. Black and Hispanic men are less likely than whites to consider housework as unfair to their wives. In contrast to couples in which children are all the biological

TABLE 4: Logistic Regression Results for Husband's Perception That Housework Is Unfair to the Wife[a]

	b	Exp {b}
Independent variables		
Controls		
Husband's age	-.043***	.958
Black couple (white couples)[b]	-1.098***	.334
Hispanic couple (white couples)	-.834**	.434
Other child-type family (all biological children)	-.445*	.641
Husband's contribution to routine housework		
Husband's perception of logged routine-hours ratio	-.265***	.767
Ideology variables		
Husband's sex-role egalitarianism	.035*	1.036
Husband's endorsement of housework sharing	-.219**	.804
Wife's human capital		
Wife's years of schooling	.111***	1.117
Equity variables		
Husband's perception of logged sporadic-hours ratio	-.345***	.708
Logged ratio of child socialization activities	-.529**	.589
Wife's frequency of child socialization activities	-.057***	.944
Husband has higher education (equal education)	.355**	1.426
Logged health-status ratio	-.562*	.570
Model χ^2	258.236***	
$\hat{R}^2$	.210	
(N = 2,109)		

[a] Only significant predictors are shown.
[b] Contrast category for multinomial qualitative predictors in parentheses.

* $p < .05$ ** $p < .01$ *** $p < .001$

children of both partners, men with adopted or foster children in the household or those in double stepfamilies are less likely to view housework as unfair to wives.

Men's contribution to routine housework has a very significant effect on their perception of fairness. Each unit increment in the logged routine-hours ratio (e.g., sharing housework equally as opposed to doing a third as much as

TABLE 5: Logistic Regression Results for Wife's Perception That Housework Is Unfair to the Wife[a]

	b	Exp{b}
Independent variables		
Controls		
Hispanic couple (white couples)	-.977***	.377
Other child-type family (all biological children)	-.542**	.582
Husband's contribution to routine housework		
Wife's perception of logged routine-hours ratio	-.223***	.800
Ideology variables		
Husband's endorsement of housework sharing	-.284***	.753
Wife's endorsement of housework sharing	.233***	1.262
Wife's human capital		
Wife's self-assessed health	-.242**	.785
Equity variables		
Wife's perception of logged sporadic-hours ratio	-.539***	.584
Logged ratio of mealtime with children	-.227**	.797
Wife's frequency of child socialization activities	-.038*	.963
Logged health-status ratio	-.555*	.574
Logged ratio of hours in paid labor	-.096*	.908
Model χ^2	229.928***	
$\hat{R}^2$	.160	
(*N* = 2,096)		

[a] Only significant predictors are shown.
[b] Contrast category for multinomial qualitative predictors in parentheses.

* p < .05 ** p < .01 *** p < .001

the wife) results in about a 23% reduction (100 [.767 - 1]) in the odds that housework is seen as unfair to the wife. It is interesting that his relative contribution is what is important: controlling for the logged routine-hours ratio, husbands' reports of how many absolute hours the wife spends in routine housework have no significant effect on fairness.

Of the ideology factors, both husbands' general sex-role egalitarianism and their endorsement of housework sharing have significant effects. The more generally egalitarian the husband, the more he sees housework as unfair to his wife. However, controlling for other ideology variables, the more that he

endorses sharing in housework, the less likely he is to report that the existing arrangement is unfair to his partner.

Of the wife's human capital factors, only her education is significant. Specifically, each additional year of schooling the wife has completed increases the odds that her husband will see housework as unfair to her by about 12%.

A number of the equity factors have significant effects. Their interpretation will be facilitated by keeping in mind that, because all logged ratios have husbands' contributions — or inputs — in the numerator of the ratio, the equity equation for husbands is the inverse of equation 2. Therefore we expect that the greater the husbands' inputs, relative to his outcomes, in areas other than routine housework, the less likely he will see housework as unfair to his wife. These predictions are borne out for inputs to sporadic housework and child socialization activities and partners' relative health. Each unit increase in the logged sporadic-hours ratio decreases the odds that he will see housework as unfair to his wife by about 29%. Similarly, the more he engages in child socialization activities and the better his health, relative to his wife, the less likely he is to see housework as unfair to her. The main effect of the wife's frequency of child socialization is additionally significant: the more often she engages in these activities, the less likely he is to see housework as unfair to her. Relative education is also significant, but in a direction contrary to equity. Husbands with more education than their wives are more likely than those with the same amount of education to see housework as unfair to their wives.

Results for Wives

The results for wives parallel those for husbands in many respects. Along with Hispanic husbands, Hispanic wives are less likely to see housework as unfair to them than are white wives. Similarly, wives with adopted or foster children in the household or those in double stepfamilies are less likely to view housework as unfair to them than are those with biological children.

Husbands' relative contribution to routine housework has the same effect in the wives' as in the husbands' equation: The greater his relative contribution, the less likely she is to see housework as unfair to her. Once again, the absolute number of hours she spends in routine housework has no effect on her evaluation of housework justice.

Of the ideology factors, only the endorsement of housework sharing by each partner has a significant effect on housework justice among wives. The more the wife endorses sharing, the more likely she is to see the existing arrangement as unfair to her. Controlling for her egalitarianism in this area, however, the more the husband endorses sharing, the less likely she is to see housework as unfair.

Wives' human capital factors have little effect on their perception of housework justice, with the exception of their self-assessed health. Yet the effect is contrary to expectation: The better a wife's health, the less likely she is to view housework as unfair to her.

Several equity factors are again significant. The husband's relative contribution to sporadic housework, mealtime with children, hours in paid employment, and relative health compared to his wife's all have the expected effects on the wife's perception of housework fairness. Each unit increment in the logged sporadic-hours ratio reduces the odds that she sees housework as unfair to her

by about 42%. Indeed, a test of the equality of coefficients for routine and sporadic housework reveals that for wives (but not husbands), husbands' contributions to sporadic housework have a significantly ($p < .002$) stronger effect on wives' perceptions of fairness than husbands' contributions to routine housework. For husbands, the men's relative contribution to child socialization reflected the impact of equity. Among wives, on the other hand, it is contributions to mealtime with children that is important. The more meals he has with the children, relative to his wife, the less likely she is to see housework as unfair. Similar to findings for husbands, the better his health compared to hers, the less likely she is to say that housework is unfair to her. Although partners' relative contributions to paid employment were not significant for husbands, the effect is significant for wives. The greater the husband's number of hours in paid labor, relative to his wife's, the less likely she is to see housework as unfair. Finally, wives' frequency of child socialization has the same effect as in the equation for husbands: The more frequently she engages in child socialization activities, the less likely she is to say that housework is unfair to her.

Discussion

Net of partners' relative contributions to the routine, or traditionally "female," work of the household, our study finds stronger support for ideology and equity than for wives' human capital as determinants of housework justice. The only index of wives' human capital having the expected effect was wives' education: Husbands of more educated wives were more likely to see housework as unfair to their partners. No such effect was observed, however, for wives. Among the latter, only a wife's health was predictive of housework fairness, but in the opposite direction to that expected: Wives in better health were less likely to say that housework was unfair to them.

The failure of factors tapping wives' general human capital to have much effect on perceptions of housework fairness for either partner was not entirely unexpected. We have argued that wives' alternatives to the marriage may have countervailing effects on judgments of fairness of the existing division of labor. On the one hand, wives with more alternatives should have greater leverage in the relationship with which to negotiate a more favorable division of housework (see England & Farkas 1986 for arguments supporting this reasoning). This means that they should be more likely to see the outcome as fair. On the other hand, their expectations for equality may well be higher, and hence they are also likely to judge any given arrangement as less fair (see Lennon & Rosenfield 1994 for an elaboration of this argument). Unfortunately, we have no information on a crucial element of wives' options: their ability to attract other mates. Although wives with more human capital may well be able to support themselves outside the marriage, most men and women do not want to spend their lives alone. Either partner's personal or physical attractiveness would be expected to exert a powerful influence on their ability to choose other partners and hence to add leverage to their negotiations in the marriage (Blumberg & Coleman 1989; England & Farkas 1986; Walster, Walster & Berscheid 1978). Indeed, South and Lloyd's (1995) recent work on marital dissolution suggests

that almost a third of recent divorces were preceded by infidelity on the part of at least one partner. They further found that, net of individual characteristics, the probability of any marriage ending in divorce is enhanced when either the husband or the wife encounters a surplus of alternative partners. The lack of any measures in the NSFH with which to assess individuals' access to alternative partners prevents further investigation of this issue with these data. Nevertheless, future research on couples' negotiations over key issues such as housework should attempt to include measures of partners' relative access to other mates. Such bargaining "chips" may provide either partner with enhanced power to achieve desirable outcomes.

Of the other factors hypothesized to affect the perception of justice in housework, ideology was significant for both husbands and wives. More egalitarian husbands tended to see housework as more unfair to their wives. The more wives endorsed the sharing of housework, the more they saw the existing arrangement as unfair to them. However, the more husbands agreed that housework should be equally shared, the less both partners saw housework as unfair to the wife. These patterns suggest that the motivation behind husbands' housework is a key factor in partners' evaluations of their division of labor. To the extent that husbands perform housework because they believe in sharing it, rather than because they have been coerced, both husbands and wives are more likely to accept this work as fairly distributed. This is consistent with literature suggesting that husbands' housework is prized by wives primarily for its symbolic value in equalizing the status of the partners (Hochschild 1989; Piña & Bengtson 1993; Ross, Mirowsky & Huber 1983). Housework performed willingly, rather than grudgingly, by husbands would necessarily have greater symbolic value to both partners.

Similarly, several variables tapping objective equity were significant. The greater the husband's relative contribution to sporadic chores and to child care, and the better his health relative to his wife's, the less likely either partner was to say that housework was unfair to the wife. Moreover, the greater the disparity between the time husbands spent working and that spent by their wives, the less likely women were to say that housework was unjustly apportioned. Of the equity factors that did not have the expected effects for either husbands or wives, education and income are especially interesting. These factors could also be said to tap spouses' relative power in the relationship. Were they to have been found significant in the expected direction, one could argue that power differentials, rather than equity, are driving the results. In contrast, it is the factors having to do with relative contributions to the labor required for the sustenance of the household that have the expected effects: sporadic housework, child care, health, and hours in paid employment. Because these factors have little to do with power differentials, the case for equity is perhaps that much stronger.

Nevertheless, there were some unexpected findings. Both partners were less likely to see housework as unfair to women the more hours the latter spent in child socialization activities. However, we see this as indicative of the fact that activities with children are a valued activity for both partners (Thompson & Walker 1989). The time that wives get to spend with children doing enjoyable activities — having private talks, taking them to recreational events, and so

forth — may very well be seen by both partners as compensating for the less interesting work of the household.

More troubling for the equity hypothesis was the fact that husbands with more education than their wives were *more* likely to see their wives as unfairly burdened with housework. This is clearly contrary to the expectation that superior education would act as an input for men that compensates for less effort in housework. Rather, it appears that status differences in favor of husbands make them more uneasy with the typical setup in which wives are shouldering most of the household chores. As the performance of housework for another's benefit has been characterized as servile (Ross, Mirowsky & Huber 1983), these husbands may be reacting to their wives being so stigmatized. This, however, suggests a much more complex interplay between outcome/input ratios and relationship dynamics than was originally hypothesized. Hochschild (1989) has already suggested such an interplay via a "balancing" strategy among wives. She observed that wives in her study with incomes superior to their husbands' balanced such a threat to their partner's status by doing *more* than their normal quota of housework. This, too, is contrary to equity principles, which predict that such wives would correspondingly *reduce* their housework load. In either case, equity considerations do not appear to operate in an affective vacuum. Perceptions of, and feelings about, the meanings that one's partner attaches to exchanges in an intimate relationship appear to intrude upon a straight calculation of one's own gain and loss. Future research needs to provide more assessments of such considerations so that the processes involved in exchanges among intimates can be better untangled.

The impact of both ideology and equity considerations in our models also suggests that equity considerations alone do not inform the perception of fairness in intimate relationships. In this sense, our findings support previous critiques of the rational choice model of negotiation (see especially Marini 1992). Exchange theorists tend to view dyadic bargaining as occurring on a level playing field, in which all partners have equal information, and in which choices are relatively unconstrained except by their potential for evoking reward (England & Farkas 1986; Kelley & Thibaut 1978; Marini 1992; Thibaut & Kelley 1959). To the extent that perceptions of justice in housework are channeled by gender-role ideology, partners may be induced by internalized cultural norms to accept arrangements that are not in their best interests. Bargaining will typically cease when each partner feels that he or she has extracted maximum profit from the exchange (Kelley & Thibaut 1978). As long as women, in particular, are taught that they have the primary obligation for housework, they will already come to the "bargaining table" with lowered expectations for "profit." Under such conditions, they will come to accept housework arrangements that are imbalanced in favor of men as "fair" (see Bielby & Bielby 1992 for a discussion of the parallel effects of ideology in constraining couples' negotiations over moving for career advancement). More work needs to be done on the ways in which negotiations over the household division of labor are constrained by partners' acceptance of such ideologies.

Some limitations of this study should be addressed in future research. First, we have assumed that there is a recursive causal relationship between the actual performance of housework and the perception of housework justice. Yet what

we have examined is the relationship between the perception of housework justice and *reports* of housework performance. It is entirely possible that these two elements, as measured, are nonrecursively linked. That is, reports of housework hours influence perceptions of justice because they reflect the actual time spent in housework. But perceptions of housework justice may also distort partners' recall of the time each spouse has actually spent doing housework. Thus, partners who see housework as unfair to the wife may recall the husband contributing fewer hours than he actually did, and vice versa. This problem, of course, is not unique to our study but rather characterizes all studies that link reported housework to its perceived fairness. Having spouses keep diaries of daily time spent in various household tasks (as, for example, in Coverman & Sheley 1986) is one way of minimizing this problem.

Second, we have touched above on the executive component of housework responsibilities. Unfortunately, the NSFH contains no information about who is responsible for the planning and supervision of housework. We were therefore not able to ascertain whether sharing the responsibility for overseeing household upkeep is a factor in judgments that housework is fairly divided. Other research suggests that women are especially troubled at being left in charge not only of doing most of the housework, but also of sensing when and if it needs to be done (Rubin 1976, 1983). As this is a major aspect of housework on which men and women are likely to differ dramatically, it should be included in future studies.

Finally, our approach to housework and child care has been limited in its disregard for individual preferences. Such a simplistic approach was dictated by the constraints of the data. The NSFH did not ask respondents the degree to which each household task or activity with children was pleasant or distasteful. Certainly the value to the partner of one's contribution to household work or child care would be reduced to the degree that it is intrinsically rewarding, and enhanced to the degree that it is not. Previously we speculated that such considerations are responsible for the finding that women's time in child socialization is unexpectedly linked to a lower likelihood that either partner will see housework as unfair to her. In making this speculation, we have assumed that both partners value time spent with their children in constructive endeavors. However, lacking measures of the relative preferences for each activity, such a speculation is, for the moment, unsupported. Scholars engaged in subsequent studies of housework fairness should ensure that assessments of individuals' preferences are included in their research designs.

Notes

1. Although some of the tasks that we have labeled sporadic, for example, shopping or driving, may be done fairly regularly, tasks 1-4 are clearly the most time-consuming tasks for wives in the sample and are therefore distinguished as the "routine" tasks. According to both partners' reports, wives spend upward of five and a half hours per week doing each of these. In contrast, shopping — the next most time-consuming chore for women — takes up only three-and-a-half hours in the typical week.

2. Other researchers have usually recoded extreme values reported for housework hours so that these sums are not above some reasonable value, such as 84 hours (Presser 1994) or 100 hours (Perry-Jenkins & Folk 1994) per week for employed persons. This has not been done here

APPENDIX: Means and Standard Deviations for All Variables in the Multivariate Analyses[a]

	Mean	Std. Dev.
Dependent variables		
Husband's perception that tasks are unfair to the wife	.288	.453
Wife's perception that tasks are unfair to the wife	.356	.479
Independent variables		
Control variables		
Duration of marriage	12.324	7.518
Cohabiting unmarried	.055	.228
Number of children	2.234	1.090
Husband's age	38.584	7.778
Black couple (white couples)[b]	.128	.334
Hispanic couple (white couples)	.072	.258
Other couple type (white couples)	.069	.254
Stepfather family (all biological children)	.216	.412
Stepmother family (all biological children)	.037	.188
Other child-type family (all biological children)	.068	.252
Husband's contribution to routine housework		
Husband's perception of logged routine-hours ratio	-1.712	1.322
Wife's perception of logged routine-hours ratio	-1.735	1.259
Husband's perception of wife's routine hours	32.027	20.976
Wife's perception of wife's routine hours	32.969	19.634
Ideology variables		
Husband's sex-role egalitarianism	.000	3.644
Wife's sex-role egalitarianism	.000	3.683
Husband's endorsement of housework sharing	3.983	.765
Wife's endorsement of housework sharing	4.246	.738

because our interest is in the impact on housework fairness of the male's *relative* contribution to housework, as assessed by the logged ratios of male to female task hours. Moreover, the males' reports of each partner's housework hours are used in the equations for males, and females' reports are used in equations for females. We assume further that the tendency to report exaggerated hours in housework will be randomly distributed across genders, and should apply equally to hours reported for either partner. For this reason, the ratios of male-to-female hours should not be biased, given that overestimates in the denominator should counterbalance overestimates in the numerator, for either gender.

3. Our definitions of routine and sporadic tasks differ somewhat from what Lennon and Rosenfield refer to as the "female" and "male" tasks. Their "female tasks" add shopping to the routine tasks enumerated above, while the "male" tasks include only paying bills, outdoor tasks, and automobile maintenance.

APPENDIX: Means and Standard Deviations for All Variables in the Multivariate Analyses[a] (Continued)

	Mean	Std. Dev.
Wife's human capital		
Wife's annual income, in thousands	10.064	11.161
Wife's years of schooling	12.835	2.590
Wife's self-assessed health	4.106	.738
Wife's weekly hours in paid labor	23.720	19.293
Equity variables		
Husband's perception of logged sporadic-hours ratio	.139	.773
Wife's perception of logged sporadic-hours ratio	.013	.765
Husband's perception of wife's sporadic hours	11.136	11.242
Wife's perception of wife's sporadic hours	11.069	10.652
Logged ratio of mealtime with children	-.210	.723
Wife's mealtime with children	9.559	3.627
Logged ratio of child socialization activities	-.151	.374
Wife's frequency of child socialization activities	16.623	4.236
Logged income ratio	1.540	1.738
Husband has higher education (equal education)	.397	.489
Wife has higher education (equal education)	.301	.459
Logged health-status ratio	.003	.299
Logged ratio of hours in paid labor	1.527	2.348
Husband has a second job	.143	.350
Wife has a second job	.069	.254

[a] Based on the unweighted data.
[b] Contrast category for multinomial qualitative predictors in parentheses.

References

Adams, J. Stacy. 1965. "Inequity in Social Exchange." Pp. 267-99 in *Advances in Experimental Social Psychology*, edited by Leonard Berkowitz. Academic Press.

Agresti, Alan. 1989. "Tutorial on Modeling Ordered Categorical Response Data." *Psychological Bulletin* 105:290-301.

____. 1990. *Categorical Data Analysis*. Wiley.

Becker, Gary S. 1991. *A Treatise on the Family*. Harvard University Press.

Benin, Mary Holland, and Joan Agostinelli. 1988. "Husbands' and Wives' Satisfaction with the Division of Labor." *Journal of Marriage and the Family* 50:349-61.

Bergen, Elizabeth. 1990. *The Multidimensional Nature of Domestic Labor: An Investigation of Husbands' Participation*. Center for Demography and Ecology, University of Wisconsin.

Bielby, William T., and Denise D. Bielby. 1992. "I Will Follow Him: Family Ties, Gender-Role Beliefs, and Reluctance to Relocate for a Better Job." *American Journal of Sociology* 97:1241-67.

Blair, Sampson Lee, and Michael P. Johnson. 1992. "Wives' Perceptions of the Fairness of the Division of Household Labor: The Intersection of Housework and Ideology." *Journal of Marriage and the Family* 54:570-81.

Blumberg, Rae L., and Marion T. Coleman. 1989. "A Theoretical Look at the Balance of Power in the American Couple." *Journal of Family Issues* 10:225-50.

Cook, Karen S., and Karen A. Hegtvedt. 1983. "Distributive Justice, Equity, and Equality." *Annual Review of Sociology* 9:217-41.

Coverman, Shelley, and Joseph F. Sheley. 1986. "Change in Men's Housework and Childcare Time, 1965-1975." *Journal of Marriage and the Family* 48:413-22.

England, Paula, and George Farkas. 1986. *Households, Employment, and Gender: A Social, Economic, and Demographic View.* Aldine de Gruyter.

Farkas, Arthur J., and Norman H. Anderson. 1979. "Multidimensional Input in Equity Theory." *Journal of Personality and Social Psychology* 37:879-96.

Glass, Jennifer, and Tetsushi Fujimoto. 1994. "Housework, Paid Work, and Depression among Husbands and Wives." *Journal of Health and Social Behavior* 35:179-91.

Hartmann, Heidi I. 1981. "The Family as the Locus of Gender, Class, and Political Struggle: The Example of Housework." *Signs: Journal of Women in Culture and Society* 6:366-94.

Hochschild, Arlie. 1989. *The Second Shift: Working Parents and the Revolution at Home.* Viking Penguin.

Idler, Ellen L., and Ronald J. Angel. 1990. "Self-Rated Health and Mortality in the NHANES-I Epidemiologic Follow-up Study." *American Journal of Public Health* 80:446-52.

Jasso, Guillermina. 1978. "On the Justice of Earnings: A New Specification of the Justice Evaluation Function." *American Journal of Sociology* 83:1398-1419.

Kelley, Harold H., and John W. Thibaut. 1978. *Interpersonal Relations: A Theory of Interdependence.* Wiley.

Lennon, Mary C., and Sarah Rosenfield. 1994. "Relative Fairness and the Division of Housework: The Importance of Options." *American Journal of Sociology* 100:506-31.

Marini, Margaret M. 1992. "The Role of Models of Purposive Action in Sociology." Pp. 21-48 in *Rational Choice Theory: Advocacy and Critique,* edited by James S. Coleman and Thomas J. Fararo. Sage.

McKelvey, Richard D., and William Zavoina. 1975. "A Statistical Model for the Analysis of Ordinal Dependent Variables." *Journal of Mathematical Sociology* 4:103-20.

Miles, Edward W. 1987. "Selection of an Equity Formula Appropriate for Organizational Behavior Research." *Sociological Methods and Research* 15:447-66.

Morgan, Mary Y., and John Scanzoni. 1987. "Religious Orientations and Women's Expected Continuity in the Labor Force." *Journal of Marriage and the Family* 49:367-79.

Myers, Raymond H. 1986. *Classical and Modern Regression with Applications.* Duxbury Press.

Nyquist, Linda, Karla Slivken, Janet T. Spence, and Robert L. Helmreich. 1985. "Household Responsibilities in Middle-Class Couples: The Contribution of Demographic and Personality Variables." *Sex Roles* 12:15-34.

Perry-Jenkins, Maureen, and Karen Folk. 1994. "Class, Couples, and Conflict: Effects of the Division of Labor on Assessments of Marriage in Dual-Earner Families. *Journal of Marriage and the Family* 56:165-80.

Piña, Darlene L., and Vern L. Bengtson. 1993. "The Division of Household Labor and Wives' Happiness: Ideology, Employment, and Perceptions of Support." *Journal of Marriage and the Family* 55:901-12.

Presser, Harriet B. 1994. "Employment Schedules among Dual-Earner Spouses and the Division of Labor by Gender." *American Sociological Review* 59:348-64.

Ross, Catherine E. 1987. "The Division of Labor at Home." *Social Forces* 65:816-33.

Ross, Catherine E., John Mirowsky, and Joan Huber. 1983. "Dividing Work, Sharing Work, and In-Between: Marriage Patterns and Depression." *American Sociological Review* 48:809-23.

Rubin, Lillian B. 1976. *Worlds of Pain: Life in the Working-Class Family*. Basic Books.

____. 1983. *Intimate Strangers: Men and Women Together*. Harper & Row..

Sanchez, Laura. 1994. "Gender, Labor Allocations, and the Psychology of Entitlement within the Home." *Social Forces* 73:533-53.

Scanzoni, John, and Deborah D. Godwin. 1990. "Negotiation Effectiveness and Acceptable Outcomes." *Social Psychology Quarterly* 53:239-51.

Schooler, Carmi, Joanne Miller, Karen A. Miller, and Carol N. Richtand. 1984. "Work for the Household: Its Nature and Consequences for Husbands and Wives." *American Journal of Sociology* 90:97-124.

Shelton, Beth Anne. 1990. "The Distribution of Household Tasks: Does Wife's Employment Status Make a Difference?" *Journal of Family Issues* 11:115-35.

South, Scott J., and Glenna Spitze. 1994. "Housework in Marital and Nonmarital Households." *American Sociological Review* 59:327-47.

South, Scott J., and Kim M. Lloyd. 1995. "Spousal Alternatives and Marital Dissolution." *American Sociological Review* 60:21-35.

Spitze, Glenna. 1988. "Women's Employment and Family Relations: A Review." *Journal of Marriage and the Family* 50:595-618.

Sprecher, Susan. 1986. "The Relation between Inequity and Emotions in Close Relationships." *Social Psychology Quarterly* 49:309-21.

Sweet, James, Larry Bumpass, and Vaughn Call. 1988. *The Design and Content of the National Survey of Families and Households*. University of Wisconsin, Center for Demography and Ecology.

Szinovacz, Maximiliane, and Paula Harpster. 1993. "Employment Status, Gender Role Attitudes, and Marital Dependence in Later Life." *Journal of Marriage and the Family* 55:927-40.

Thibaut, John W., and Harold H. Kelley. 1959. *The Social Psychology of Groups*. Wiley.

Thompson, Linda. 1991. "Family Work: Women's Sense of Fairness." *Journal of Family Issues* 12:181-96.

Thompson, Linda, and Alexis J. Walker. 1989. "Gender in Families: Women and Men in Marriage, Work, and Parenthood." *Journal of Marriage and the Family* 51:845-71.

Walster, Elaine, G. William Walster, and Ellen Berscheid. 1978. *Equity: Theory and Research*. Allyn and Bacon.

Ward, Russell A. 1993. "Marital Happiness and Household Equity in Later Life." *Journal of Marriage and the Family* 55:427-38.

Physical Impairment and the Diminishing Effects of Income*

JOHN MIROWSKY, *Ohio State University*
PAUL NONGZHUANG HU, *Ohio State University*

Abstract

The effect of income on physical impairment steepens at lower levels of income. Why? Two national surveys show the following. Basic needs: the level of impairment and its rate of increase rise sharply as household income drops below the 20th percentile; above that level greater income has little effect; economic hardship explains much of the pattern. Resource substitution: education reduces the association between impairment and low income, but low education often coincides with low income. Compounding problems: impairment suppresses the rise of income, concentrating economic and health problems over time. Status lifestyle: Exercise increases with income but explains little of the association with impairment. Social background: stable traits that affect both income and impairment account for about half of their association, with education the largest factor.

Previous research shows that income is associated with a greater sense of health and well-being and with lower psychophysiological distress, lower functional impairment, and lower risks for morbidity and mortality (Broom 1984; Gold & Franks 1990; Haghighatian & Baer 1990; Hay 1988; Hirdes et al. 1986; Maddox & Clark 1992; Mirowsky & Ross 1989; Pappas et al. 1993; Rogers 1992; Williams 1990). Income provides the means to improve health through better living conditions, more accessible and higher quality medical care, and reduced stress and anxiety. Less clear is the exact form of the relationship between income and health. Most studies implicitly assume that the association is linear: a unit

* *We thank Catherine E. Ross for use of the data from her survey of Work, Family and Well-Being that was funded by the National Science Foundation (SES-8916154). Sampling, pretesting, and interviewing were conducted by the Survey Research Laboratory of the University of Illinois. Data for the National Survey of Personal Health Practices and Consequences were collected by the National Center for Health Statistics (NCHS), U.S. Department of Health and Human Services. They were made available, in part, by the Inter-University Consortium for Political and Social Research (ICPSR). Mirowsky and Hu are equal coauthors of this report.* *Address communications to John Mirowsky, Department of Sociology, Ohio State University, 300 Bricker Hall, 190 North Oval mall, Columbus, Ohio 43210-1353 or e-mail: mirowsky.1@osu.edu.*

 Social Forces, March 1996, 74(3):1073-1096

increase in income produces a unit improvement in health. Recent evidence challenges this assumption. Differences or changes in income may influence health more strongly among those with marginal or modest incomes than among those with affluent ones. House et al. (1990) find only slight health differences associated with greater income above $20,000 but markedly worse health as incomes drop below $20,000 (at or near the poverty levels). Diener et al. (1993) find a similar nonlinear effect of income on subjective well-being using U.S. national data. Epelbaum's (1990) review of "sociomonetary relationships" between measures of income or wealth and measures such as rates of morbidity, mortality, and health finds curvilinear patterns that suggest diminishing effects of greater income. Such findings strongly suggest that the effect of additional income on health diminish as the levels of income rise.

As Angell (1993) points out, at this stage we do not know with any certainty that the protective effects of higher income apply at the upper levels of privilege. We pose three questions. First, is there really a diminishing effect of greater income on health? Second, is there a threshold below which greater income improves health and above which it does not? Third, what, if anything, causes the effect of income to diminish at the higher income levels?

Although a number of findings suggest diminishing effects of greater income on health (e.g., Epelbaum 1990; House et al. 1990), relatively few suggest why the effect should taper off at higher levels. The most obvious explanation refers to the fundamental reason why income might effect health in the first place. The *basic-needs hypothesis* says income improves health because it helps individuals meet universal human needs (Williams 1990). Once needs are met, greater income can have no more effect. Indeed, one may define poverty as the lack of means necessary for physical well-being — as the inability to meet basic needs for food, shelter, clothes, and care. The higher one's income, the more likely these needs will be met. At some level of income most people can afford adequate food, housing, and clothes. Reasonably good medical care might require somewhat higher income than other needs, but at some level of income most people have what they need. The effect of greater income on health diminishes as it rises beyond the amount needed to provide the essentials. Thus, the basic-needs hypothesis implies two things. First, it implies the existence of a critical point — a "poverty line" defined in terms of health. Below the line greater income improves health; above the line it does not. Second, it implies that economic hardship mediates a large fraction of the association between low income and poor health: In the absence of difficulty paying bills and buying necessities like food, clothes, or medicine, low income does not translate into poor health.

The *resource-substitution hypothesis* supplements the needs hypotheses. It says that income is only one of the resources that help people meet their health requirements. Greater income has more impact in the absence of other resources, and, conversely, less impact in their presence. All the useful resources tend to be scarce at the bottom of the socioeconomic ladder and abundant at the top. The scarcity of other resources makes the health of low-income people more sensitive to increases in income. The abundance of other resources makes the health of high-income people less sensitive to increases in income.

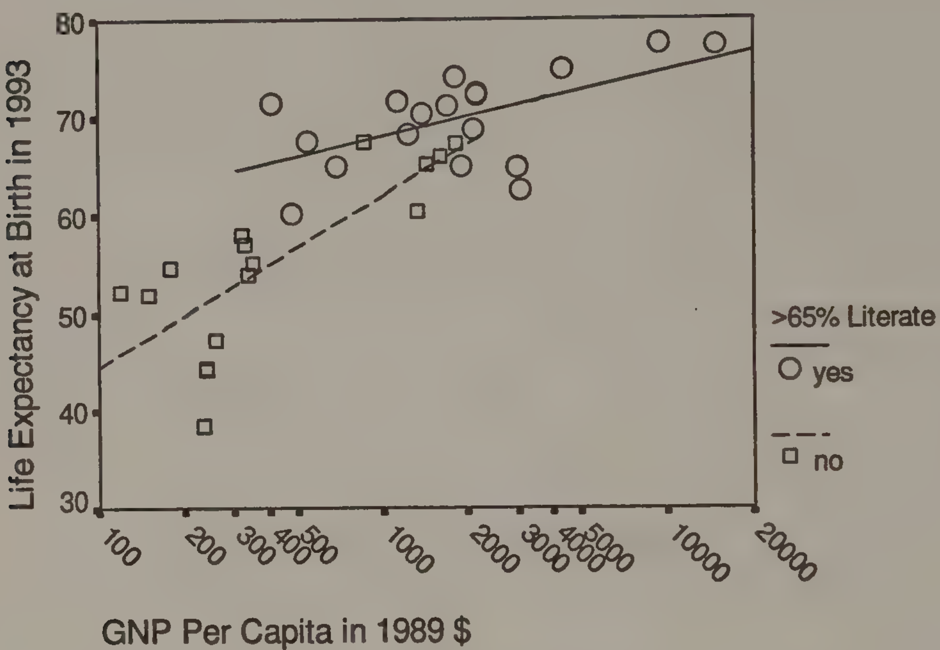
Cross-national data on life expectancy at birth suggests that education may operate as an alternative resource for health (Mosley & Cowley 1991; Sen 1993). A life expectancy of 65 years or more generally coincides with a Gross National Product (GNP) per capita above $2,000. However some countries with low GNP per capita but relatively high literacy rates also have life expectancy of 65 years or more. Sri Lanka stands out among the countries in this category. With an adult literacy rate of 87.1%, Sri Lanka has a life expectancy of 71.5 years despite having less than $500 of GNP per capita (U.S. Bureau of the Census 1993). Life expectancy and GNP come from the U.S. Bureau of the Census International Data Base (U.S. Bureau of the Census 1993). Literacy rates come from the United Nations Educational, Scientific and Cultural Organization (Hoffman 1991). Figure 1 plots the life expectancy at birth by GNP per capita for 34 countries with populations of 5 million or more that report literacy rates. The figure also shows the OLS regression lines for countries with low and high literacy (below and above the 65% level). Predicted life expectancy increases more rapidly with GNP for the lower-literacy countries than for the higher-literacy ones. GNP and literacy appear to act as alternative inputs to life expectancy.

U.S. survey data also suggests that education and income may substitute for each other as alternative resources. Ross and Huber (1985) find that income has a weaker negative association with economic hardship at higher levels of education. Low income raises the average frequency of difficulty paying for needed food, clothing, and medical care, but *less* so at successively higher levels of education. Higher income reduces the risk of economic hardship the most for the poorly educated. Presumably the better educated know how to meet their household needs with less income. Of course, higher income often implies higher education. Implicitly, then, an increment in income reduces economic hardship less among the prosperous than among the poor, because the prosperous tend to be better educated. In essence, higher education makes each additional dollar less necessary for meeting needs. Because of the positive correlation between education and income, additional dollars reduce economic hardship less at higher levels of income. The same may be true for health outcomes.

The basic-needs and resource-substitution hypotheses focus on probable mediators and moderators of income's effect on health, but they ignore the possibility of dynamic feedback. The *compounding-problems hypothesis* argues that processes of social selection based on health concentrate the people with poorest health at the lowest levels of income. Poor health reduces income and low income reduces health. Regardless of which problem begins the cycle, one precipitates the other. Low income increases the risk of health problems that simultaneously magnify economic hardship and hamper efforts to escape poverty. Over a period of time, low initial income increases the rate at which health problems emerge; and initial health problems reduce the rate at which income rises. Over a series of periods the problems compound, concentrating poor health among those with low income and low income among those with poor health. Thus, the compounding of economic and health problems might create an increasing gradient of poor health at lower and lower levels of income.

The *status-lifestyle hypothesis* suggests another explanation of income's diminishing effect on health. People need physical activity in order to keep fit.

FIGURE 1: Life Expectancy by GNP and Literacy



In the U.S. today, working, commuting, and shopping require little physical effort, making exercise a leisure activity. Among U.S. adults the average amount of weekly exercise increases with socioeconomic status — particularly with education but also with income (Hayes & Ross 1986). Perhaps a daunting environment and the struggle to make ends meet make it hard for many lower-income people to exercises regularly. Perhaps exercise operates as a modern form of conspicuous consumption, or as a public display of proper and prestigious attitudes. Or perhaps exercise reflects an orientation toward the future — a biological savings plan with sacrifices made today for benefits expected in the future. Whatever the reason for the positive correlation between income and exercise, the correlation may contribute to the concentration of poor health at the lowest levels of income.

A final hypothesis notes that income and health may have common social precursors that create correlation between them. The *social-background hypothesis* says that sex, race, marriage, and education account for much of the association between income and health. Females, nonwhites, the unmarried and the relatively uneducated tend to have both low income and poor health. To the extent that these aspects of social background have nonlinear effects on either income or health, they could produce a correlation that is stronger at low levels of income than at high ones. To take a simple example, suppose that racial differences in health and income account for the correlation. Blacks have more health problems than whites; the percentage black increases sharply toward the bottom of the income distribution; thus the level of health problems rises sharply toward the bottom of the income distribution. As always, the possibility

of spurious association needs to be examined. In this case, though, one should not think of the spurious component as a false and irrelevant correlation to be excised and then ignored. Suppose that social background accounts for a substantial fraction of the total correlation between income and health. If so, the finding would reveal part of the means by which poor health becomes concentrated among persons with low income.

SUMMARY

Does the incremental effect of income on health diminish at higher levels of income? And if so, why? We will examine the profile of the relationship and test several hypothesis, using physical impairment as our index of health (see the measurement section for details):

(1) The *basic-needs hypothesis* says that higher income predicts lower impairment up to a point but not beyond *and* that adjustment for economic hardship substantially and significantly reduces the negative association between income and impairment;

(2) The *resource-substitution hypothesis* says that education reduces the strength of the association between income and physical impairment such that increments in income have diminishing effects at higher levels of income because of the coinciding higher levels of education;

(3) The *compounding-problems hypothesis* says that income reduces the rise of impairment over time *and* that impairment reduces the rise of income over time;

(4) The *status-lifestyle hypothesis* says that adjusting for exercise substantially and significantly reduces the nonlinear association between income and impairment, and;

(5) The *social-background hypothesis* says that adjusting for sex, race, marriage, and education substantially and significantly reduces the nonlinear association between income and impairment.

Method

SAMPLES

We use two data sets to enhance the generalizability of the results and to compare cross-sectional with follow-up results. The data come from two telephone surveys of national probability samples of U.S. households. We chose the data sets because each contained adequate and comparable measures of impairment, income, economic hardship, and exercise, as well as measures of education, sex, race, marriage, and age. Each survey has its strengths and weaknesses, with the strengths of one covering the weaknesses of the other. The Work, Family, and Well-Being (WFW) survey has particularly good measures of income and economic hardship (described below), but the data are cross-sectional, and the index of exercise has only two items. The Health Practices (HP) survey has weaker measures of income and economic hardship than the WFW survey, but it has follow-up interviews and a more reliable index of exercise.

The 1990 U.S. Survey of Work, Family, and Well-Being (WFW) used random-digit dialing to ensure the inclusion of unlisted numbers, calling each selected number up to ten times (Waksberg 1978). Within each household the person eighteen years old or older with the most recent birthday was selected as respondent. (This randomly selects a household respondent [O'Rourke & Blair 1983].) Up to ten attempts were made to speak with the selected respondent and complete the interview. Eighty-two percent of contacted households produced completed interviews, yielding a total of 2,031 respondents ranging in age from 18 to 90.

The National Survey of Personal Health Practices and Consequences (HP) was taken in 1979 with reinterviews in 1980. The sampling began with a random selection of telephone exchanges, followed by a random selection of telephone households in proportion to the number of households served by each exchange, followed by the random selection of an eligible respondent from each household. The sample was limited to adults older than 17 and younger than 65 at the time of the first interview. Eighty-one percent of the selected phone numbers that were working, residential, and owned by a household with at least one eligible respondent produced a complete interview in wave 1. A total of 3,025 respondents were interviewed the first year and 2,436 were reinterviewed a year later.

MEASURES

The two surveys provide roughly comparable measures, as described below.

Impairment

Our analyses measure health in terms of impaired physical functioning using indexes of standard content described below. Functional impairments such as difficulty grasping, seeing, or using stairs correlate with other measures of health such as subjective health, work disability, dependence on others, diagnosed medical conditions (like heart disease or diabetes), and inability to perform common daily activities (like shopping or cooking) (Johnson & Wolinsky 1993; Nagi 1976; Waldron & Jacobs 1988). Physical impairment detracts from the subjective quality of life, interferes with many types of employment, and lowers chances for survival. As a measure of health, impairment provides certain advantages compared to others: it is less subjective than perceived health; it is less affected by access to medical care than counts of diagnosed conditions, physician visits, hospital stays, and the like; it is less episodic than symptoms such as headaches; and it is relevant to a broader population than measures geared to hospitals and nursing homes (McDowell & Newell 1987).

The Work, Family and Well-Being (WFW) index of physical impairment has seven questions asking the respondents to state the amount of difficulty they have with (1) going up and down stairs; (2) kneeling or stooping; (3) lifting or carrying objects less than 10 pounds, like a bag of groceries; (4) using hands or fingers; (5) seeing, even with glasses; (6) hearing; and (7) walking. Those who say they have no difficulty are coded 0; those who say they have some difficulty

are coded 1; and those who say they have a great deal of difficulty are coded 2. The index averages the seven responses. Alpha reliability is .80.

The Health Practices index of physical impairment has five questions asking the respondents to state the amount of difficulty they have with (1) walking; (2) using stairs or inclines; (3) standing or sitting for long periods; (4) using fingers to grasp or handle; (5) lifting or carrying something as heavy as 10 pounds. Those who say they have no difficulty are coded 0; those who say they have some difficulty are coded 1; and those who say they have a great deal of difficulty are coded 2. The index averages the responses to all five questions. The alpha reliability is .84 for the first wave and .82 for the second.

Income

The 1990 WFW survey asked, "During 1989, what was your *total household income,* including income from all members of the household and from all sources, before taxes?" Anyone who did not report an amount was asked, "Could you tell me, was it less than $30,000 or more than $30,000?" If they said it was less, they were asked,"Was it less than $20,000?" The probes continued until the person's income could be placed in one of the following ranges (in thousands): < 10, 10-20, 20-30, 30-40, 40-50, 50-75, 75-100, and > 100. The respective ranges were assigned the following income values (in thousands of dollars): 5, 15, 25, 35, 45, 62, 87, and 100. If someone answered "Don't know" or refused at any point in the probe, they were coded as missing. Of the 2,031 persons in the sample, 1,243 reported an exact amount on the initial question (61.2%), 492 cooperated with the probe and have an approximate value (24.2%), and 296 refused or claimed not to know and have missing values (14.6%).

The regression analyses deflate the 1989 incomes in the 1990 WFW survey to 1978-dollar equivalents for comparison to the 1979 Health Practices survey. To convert, we multiplied income in the 1990 WFW data by the ratio of 1989 consumer price index and 1978 consumer price index (.766/1.380 = .555).

The 1979 Health Practices survey asked, "Please tell me into which of the following groups your family's combined income fell in 1978, before taxes: Less than $5,000, $5,000 to $9,999, $10,000 to $14,999, $15,000 to $24,999, or $25,000 or more?" A year later they were asked the same question about income in 1979. Responses are coded respectively (in thousands of dollars) as 2.5, 7.5, 12.5, 20, or 30. Of those interviewed, 93.0% reported their 1978 income and 95.2% reported their 1979 income. The rest have missing values.

Economic Hardship

In the WFW survey economic hardship is measured as the mean response to the three questions: "During the past twelve months, how often did it happen that you (a) did not have enough money to buy food, clothes, or other things your household needed? (b) did not have enough money to pay for medical care? (c) had trouble paying the bills? Would you say never (coded 0), not very often (1), fairly often (2), or very often (3)?" The hardship index averages the responses to all three questions. Its alpha reliability is .82.

In the 1979-80 Health Practices survey economic hardship is measured by two questions, "Does it ever happen that you do not have enough money to

afford the kind of medical care you or your family should have" (coded no = 0 and yes = 1), and "Please tell me if a serious financial difficulty or problem happened in your life during the past five years" (no = 0, yes = 1). The hardship index is the mean response to the two questions. Reliability is .58 in the first wave and .67 in the second.

Exercise

In the WFW survey the exercise index is the mean of two individual measures: *walking* and *strenuous* exercise. *Walking* is measured as the number of days walked per week. Respondents were asked, "How often do you take a walk? Would you say never (coded 0), once a month or less (.25), about twice a month (.5), about once a week (1), twice a week (2), three times a week (3), more than three times a week (5), or every day (7)?" Strenuous exercise is measured by the question "How often do you do strenuous exercise such as running, basketball, aerobics, tennis, swimming, biking, and so on?" Responses are coded as the number of days per week in the same way as *walking*, except that the highest response category are those who exercise more than three times a week (coded 5). Because of a small correlation between the weekly frequency of strenuous exercise and that of walking, the reliability of the two-item index is only .27.

In the Health Practices survey, *exercise* is measured by seven questions concerning leisure time physical activities: "Please tell me how often you participate in these activities. How often do you (a) go swimming in the summer; (b) take long walks; (c)work on a physically active hobby such as dancing or gardening; (d) go jogging or running; (e) ride a bicycle; (f) do calisthenics or physical exercise; (g) participate in any active sports I haven't already mentioned?" Responses are coded never (1), rarely (2), sometimes (3), often (4). The index is the mean response to the seven questions. A high score on the index indicates higher levels of exercise. The alpha reliability for this index is .74 in both waves.

Education

The 1990 WFW survey asked, "What is the highest grade or year in school that you have completed?" Responses are coded in years up through 16 (for a college degree), with some graduate school experience assigned a score of 17 and a graduate or professional degree assigned a score of 18. The 1979 Health Practices survey asked, "What was the last grade or year of school you completed?" The response categories are none, 1-4 years, 5-6 years, 7-8 years, 9-11 years, 12 years (completed high school or equivalent), 13-15 years (some college or trade school), and 16 or more years (completed college or more). The respective codes are 0, 2.5, 5.5, 7.5, 10, 12, 14, and 16.

Other Social Background

The regressions must be adjusted for sociodemographic variables that might create a spurious association between income and impairment. The candidates include variables with moderate to large effects on both household income and health: age, sex, race, and marital status. Both surveys measure *age* in number of years and define dummy variables contrasting *females* (coded 1) with males

(0), *whites* (1) with nonwhites (0), and persons married or living with someone as *married* (1) with those who are single, divorced, separated, or widowed (0).

ANALYSES

Our analyses fall into three sets. The first set describes the cross-sectional profile of impairment by level of income. It shows whether the slope flattens at high levels of income, shows whether there is a level of income below which the slope is negative and above which it is not, and specifies a functional form that describes the curve. The second set reports cross-sectional regressions that bear on four of the hypotheses. Beginning with a bivariate regression, we progressively add adjustments for social background (precursors of both income and impairment), economic hardship, exercise, and the interaction of education with income. We run the same set of regressions using both data sets. Cross-sectional regressions leave some ambiguity about causal direction. Nevertheless, they provide information that can support or contradict specific hypotheses. Cross-sectional regressions also provide a better test of the social-background hypothesis than can over-time regressions because an ascribed status such as race or sex, or a stable one such as education or marriage, may determine the initial level of income and impairment but not the amount of change in either one net of those initial levels. The third and final set of analyses specifies over-time regressions that allow us to rule out pure selection (health causes income rather than income causing health) and to investigate the possibility of compounding problems (health and income affect each other).

Results

CROSS-SECTIONAL PROFILE

The basic association between income and physical impairment is negative. The profile of the association is consistent with the idea that income reduces impairment by helping to meet basic needs. Figure 2 shows the crude mean impairment scores by $10,000 income groups in the 1990 WFW data. The bold line profiles the mean impairment across levels of income. The dotted lines show the 95% confidence intervals for the means. A one-way analysis of variance finds both the linear trend and the deviation from linearity significant at $p < .001$ ($r = -.209$, $eta = .304$). The profile of means in the graph makes the nonlinearity apparent. Mean levels of impairment drop the most in the step from the lowest category of income to the second lowest — ranges with upper bounds at roughly the 6th percentile and the 20th. The adjacent confidence intervals do not overlap. Mean impairment drops next most in the step from the second category of income to the third, and again the adjacent confidence intervals do not overlap. From that boundary the mean seems to slope down gently through the third, fourth, and fifth categories of income, which have overlapping confidence intervals. Taken together, those categories of income range from the 20th percentile to the 76th. The third largest drop in mean impairment comes at the step from the fifth to the sixth category, and again the adjacent confidence intervals do not overlap. From the top of that category at

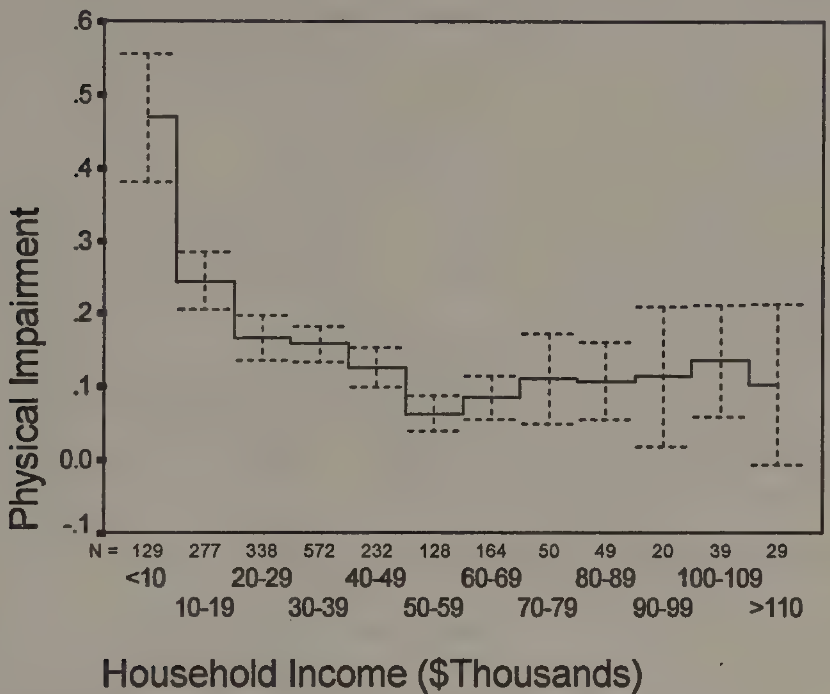
about the 83rd percentile to the top of the income distribution at the 100th percentile impairment seems to rise gently but the confidence intervals overlap considerably, meaning that the apparent rise is not statistically significant. Most of it probably reflects the positive association between age and income.

The earlier 1979-80 Health Practices data also show a diminishing negative association between impairment and income. The Health Practices survey's measure of income is crude compared to that in the WFW survey, but the overall pattern remains broadly similar. The steepest drop comes between the two lowest categories of income, which have their upper bounds at the 6th and 18th percentiles. The full bars in Figure 3 represent the mean 1980 impairment. The top part of each bar represents the mean impairment in 1979, and the bottom part represents the average increase over the one-year follow-up. One-way analyses of variance shows that both the linear trend and the deviation from linearity are significant at $p < .001$ in both years ($r = -.167$ and $-.173$ and $eta = .215$ and $.231$ in 1979 and 1980 respectively). The one-year increase in impairment is significantly greater than zero at $p < .001$ for all categories of income, despite the fact that the sample was limited to adults younger than 65. The linear trend in one-year change and its deviation from linearity are not significant in an analysis of variance ($p = .281$ and $.234$ respectively), but the correlation becomes significant after taking the cube root of income, as described below.

Both surveys show the greatest difference in mean levels of physical impairment between the lowest category of income and the next highest one. Both show diminishing steps down in average physical impairment in successively higher categories of income. Both the 1990 cross-section and the changes from 1979 to 1980 suggest a slight but statistically insignificant rebound at high levels of income, possibly because income and impairment both increase with age. Thus, regressions should adjust for age.

The curvilinear relationship between impairment and income can be approximated by transformations that shrink successive intervals, mathematically mimicking the diminishing effects. The transformations that represent diminishing effects also correct income's highly skewed distribution. Figure 2 shows the number of cases in each $10,000 bracket of income in the 1990 WFW data. The number of cases increases progressively up to the fourth bracket and decreases progressively in the next eight. The skewed distribution suggests that income may have diminishing effects at high levels partly because larger and larger increases are required to move up a rank. Because a positive skew can be reduced by taking a logarithm, a square root, or a cube root (Hamilton 1992), taking the logarithm in the 1990 WFW data reverses the skewness from 3.363 to -.885. Taking the square root brings it down to .668 and taking the cube root brings it down to .057. The standard error of the skewness is .054 regardless of the transformation, so the t values are 62.278, -16.389, 12.741, and .688 respectively. Thus the skewness of income in its cube-root form is not significant ($t < 1.960$). The cube-root transformation of income also makes its bivariate correlation with change in impairment become statistically significant. The transformation increases the t value of the correlation from -1.273 to -1.804, which lowers the one-tailed p value from .102 to .036 (or from .203 to .071 two-

FIGURE 2: Physical Impairment by Level of Income, 1990 WFW Data[a]



[a] The solid line profiles the means for impairment by level of income; the dashed lines represent the 95% confidence intervals around those means.

tailed). The regressions that follow use the cube-root of income in thousands of dollars ($\$k^{1/3}$).

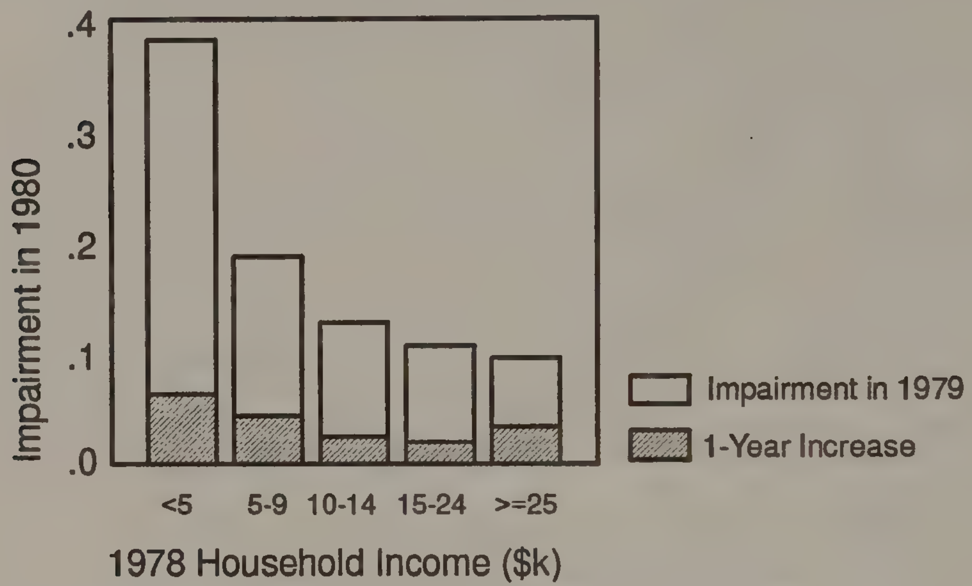
CROSS-SECTIONAL REGRESSIONS

The cross-sectional regressions provide information relevant to the hypotheses of basic needs, resource substitution, status lifestyle, and social background. Table 1 shows the cross-sectional regressions for the 1990 WFW data, and Table 2 shows parallel cross-sectional regressions in the 1979 Health Practices data.

Basic Needs

A substantial fraction of income's estimated total effect on impairment appears attributable to economic hardship. In the terms of causal analysis, the coefficient of income adjusted for its precursors represents its "total causal effect" in the model (Davis 1985). Step 2 in Tables 1 and 2 shows the regression coefficient with adjustment for social background and step 3 adds adjustment for economic hardship. The adjustment for economic hardship reduces income's coefficient by an additional 32.7% in the WFW data ([.052 - .035]/.052 = .327) and by 50.0% in the Health Practices data ([.094 - .047]/.094 = .500). The reduction is 1.545

FIGURE 3: Impairment by Income, 1979 and 1980 Health Practices Data



standard errors in the WFW data and by 3.615 standard errors in the Health Practices data. Taken together these results seem consistent with the argument that economic hardship accounts for part of the association between low income and physical impairment.

Resource Substitution

Both sets of regressions show that education significantly reduces the relationship between income and impairment. The interaction term is significant and positive in step 5 of Tables 1 and 2. In each case the coefficient of income indicates that it has a significant negative effect on impairment at the sample's mean level of education. (Education is measured as a deviation from its mean, so the coefficient of income in the interaction model represents its effect at the sample's mean level of education [Aiken & West 1991]). The positive coefficient for the interaction of income by education indicates that income's negative effect gets smaller as education rises. The left panel of Figure 4 illustrates the prediction for 10, 12, and 14 years of education in the 1990 WFW data, with the 1989 dollars converted to 1978 equivalents (for comparison to the Health Practices cross-sectional and over-time results). The cross-sectional regressions based on the 1979 Health Practices data would show a similar pattern. Both data sets yield a significant interaction between income and education. The negative association between income and impairment becomes stronger at lower levels of education. Thus, consistent with the alternative-resource hypothesis, the combination of low education with low income sharpens the association between low income and physical impairment.

TABLE 1 Physical Impairment Regressed on (1) Income, Progressively Adjusting for (2) Age, Sex, Marital Status, Race, and Education; (3) Economic Hardship; (4) Exercise and; (5) the Interaction of Education and Income[a]

	Step 1		Step 2		Step 3		Step 4		Step 5	
	b	β	b	β	b	β	b	β	b	β
Income[b]	-.127*** (.011)	-.250	-.052*** (.011)	-.102	-.035*** (.012)	-.068	-.033** (.011)	-.065	-.038*** (.011)	-.075
Age[b]			.006*** (.000)	.352	.007*** (.000)	.382	.007*** (.000)	.373	.006*** (.000)	.351
Female			.027* (.012)	.044	.019 (.012)	.030	.011 (.012)	.018	.014 (.012)	.022
White			-.014 (.018)	-.015	-.004 (.018)	-.005	-.007 (.018)	-.007	-.002 (.017)	-.003
Education[c]			-.022*** (.002)	-.190	-.020*** (.002)	-.173	-.019*** (.002)	-.165	-.073*** (.009)	-.629
Married			-.045*** (.013)	-.074	-.045*** (.013)	-.074	-.048*** (.012)	-.078	-.047*** (.012)	-.077
Hardship					.057*** (.009)	.132	.056*** (.009)	.127	.053*** (.009)	.122
Exercise							-.009*** (.002)	-.113	-.009*** (.002)	-.112
Income × Education									.021*** (.003)	.477
Intercept	.504		.331		.256		.306		.300	
R^2	.063		.234		.249		.261		.275	
(N = 2,020)										

[a] 1990 Work, Family, and Well-being data. Standard errors are in parentheses.
[b] Income is measured as the cubed root of dollars in thousands. Missing cases are assigned the mean income. The regressions adjust for a dummy variable representing the cases assigned the mean income (missing = 1, else = 0) (Cohen 1968).
[c] Education and age are measured as deviations from their respective means.

† $p < .10$ (two-tailed) * $p < .05$ ** $p < .01$ *** $p < .001$

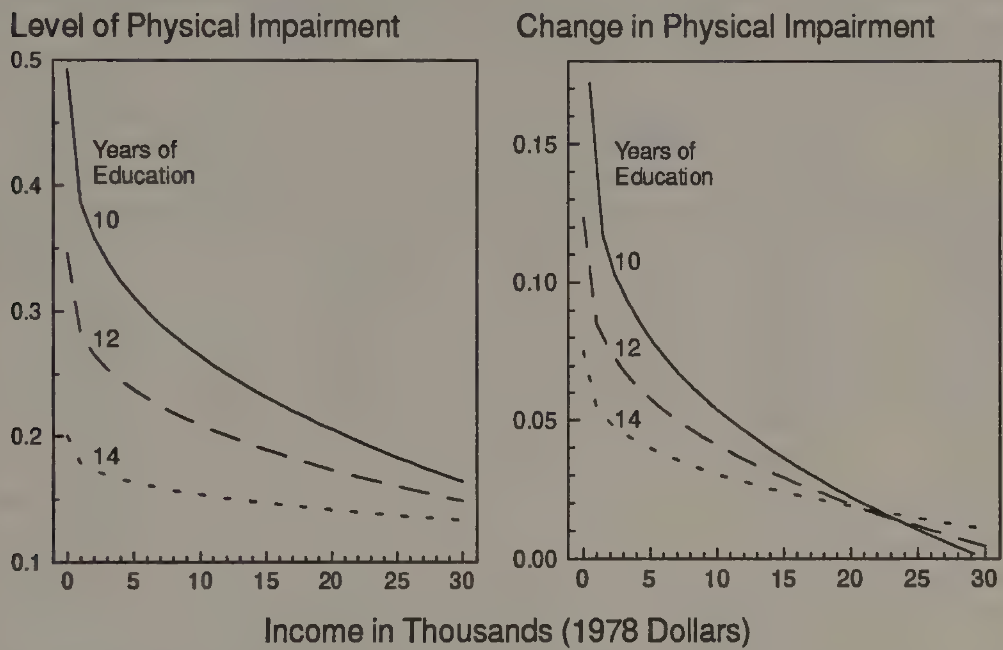
TABLE 2: Physical Impairment Regressed on (1) Income, Progressively Adjusting for (2) Age, Sex, Marital Status, Race, and Levels of Education; (3) Economic Hardship; (4) Exercise; and (5) the Interaction of Education and Income[a]

	Step 1		Step 2		Step 3		Step 4		Step 5	
	b	β	b	β	b	β	b	β	b	β
Income[b]	-.141***	-.221	-.094***	-.148	-.047***	-.074	-.031**	-.049	-.024***	-.038
	(.011)		(.013)		(.013)		(.013)		(.012)	
Age[c]			.005***	.210	.006***	.233	.004***	.147	.003***	.128
			(.000)		(.000)		(.000)		(.000)	
Female			.041***	.062	.040***	.062	.033***	.050	.035***	.054
			(.011)		(.011)		(.011)		(.011)	
White			-.042**	-.047	-.025	-.028	-.017	-.019	-.009	-.010
			(.016)		(.015)		(.015)		(.015)	
Education[c]			-.035***	-.133	-.031***	-.118	-.021***	-.081	-.186***	-.704
			(.005)		(.005)		(.005)		(.020)	
Married			-.013	-.019	-.012	-.018	-.016	-.023	-.015	-.023
			(.013)		(.012)		(.012)		(.012)	
Hardship					.209***	.216	.201***	.208	.198***	.205
					(.017)		(.017)		(.017)	
Exercise							-.016***	-.223	-.016***	-.223
							(.001)		(.001)	
Income × Education									.068***	-.632
									(.008)	
Intercept	.467		.374		.197		.428		.397	
R^2	.053		.130		.169		.205		.224	
(N=3,010)										

[a] Health Practices data. Standard errors are in parentheses.
[b] Income is measured as the cubed root of dollars in thousands, weighted to convert to 1979 dollar value. Missing cases are assigned the mean income. The regressions adjust for a dummy variable representing the cases assigned the mean (missing = 1, else = 0) (Cohen 1968).
[c] Education and age are measured as deviations from their respective means.

+ p < .10 (two-tailed) * p < .05 ** p < .01 *** p < .001

FIGURE 4: Impairment and Its Change Over One Year Predicted from Income and Education[a]



[a] The cross-sectional model on the left uses the 1990 WFW data. The over-time model on the right uses the 1979-80 Health Practices data.

Status Lifestyle

Habits of exercise associated with higher income appear to account for little or none of the negative cross-sectional association between income and impairment. Exercise has the predicted negative coefficient in both samples (although, given the cross-sectional nature of the data, infrequent exercise might *result* from impairment). Nevertheless, adjustment for exercise changes income's coefficient very little, as shown by comparison of steps 3 and 4 in Tables 1 and 2. The biggest change is in the Health Practices data (which has a more reliable exercise index), where the additional adjustment of step 4 reduces income's coefficient by 1.230 standard errors — well within the 95-percent confidence range of the coefficient in step 3. Notably, adding the adjustment for exercise in the Health Practices data reduces *education's* coefficient by 32%, or 2.000 standard errors.

Social Background

Adjustment for social background reduces income's coefficient by 59.1% in the WFW data ([.127-.052]/.127 = .591) and by 33.3% in the Health Practices data ([.141-.094]/.141 = .333), as shown by comparison of steps 1 and 2 in Tables 1 and 2. Thus, the adjustment reduces the coefficient by 6.8 standard errors in the WFW data ([.127-.052]/.011 = 6.8) and by 4.3 standard errors in the Health

TABLE 3: Changes in Physical Impairment, Income, Economic Hardship, and Exercise Regressed on Initial Impairment, Income, Hardship, Exercise, and on Age, Sex, Race, Marriage, Education, and the Interaction of Education with Initial Income and Impairment[a]

	Δ Impairment		Δ Income		Δ Hardship		Δ Exercise	
	b	β	b	β	b	β	b	β
$Impairment_{t1}$	-.318***	-.392	-.062*	-.050	.087***	.092	-.147*	-.043
	(.017)		(.025)		(.018)			(.069)
$Income_{t1}$[b]	-.033**	-.065	-.430***	-.546	-.041**	-.070	.178***	.084
	(.012)		(.017)		(.012)		(.047)	
$Hardship_{t1}$	.033*	.043	-.094***	-.079	-.564***	-.632	.002	.000
	(.016)		(.023)		(.016)		(.065)	
$Exercise_{t1}$	-.004**	-.068	.004	.015	-.000	-.002	-.337***	-.479
	(.001)		(.005)		(.004)		(.015)	
Education[c]	-.024**	-.239	-.008	-.052	-.020*	-.172	.037***	.088
	(.008)		(.013)		(.009)		(.009)	
Age[c]	.002***	.120	-.001*	-.052	-.002***	-.067	-.016***	-.200
	(.000)		(.000)		(.000)		(.002)	
Female	.021*	.041	-.014	-.018	.000	.000	.000	.000
	(.010)		(.014)		(.000)		(.040)	
White	.003	.005	.029	.025	-.023	-.026	-.052	-.017
	(.015)		(.021)		(.015)		(.059)	
Married	-.001	-.001	.117***	.142	-.003	-.005	.044	.020
	(.011)		(.016)		(.011)		(.045)	
Education x $Income_{t1}$	.009*	.214	.012*	.191	.005†	.124		
	(.004)		(.005)		(.004)			
Intercept	.163		1.087		.167		.639	
R^2	.129		.238		.345		.171	
(N = 2,427)								

[a] 1979-80 Health Practices data. Standard errors are in parentheses.
[b] Income is measured as the cubed root of dollars in thousand. Missing cases are assigned the mean income. The regressions adjust for a dummy variable representing the cases assigned the mean (missing = 1, else = 0) (Cohen 1968).
[c] Education and age are measured as deviations from their respective means.

† $p < .10$ (two-tailed) * $p < .05$ ** $p < .01$ *** $p < .001$

Practices data ([.141-.094]/.011 = 4.3). These ratios are considerably larger than the 1.96 standard errors that correspond to a probability value of .05 in a two-tailed t test. Thus, social background appears to account for a substantial and significant part of the total association between income and impairment. Education and sex appear to be the most important background contributors to the correlation. Education has a larger standardized direct effect than any background measure other than age. Age acts as a suppressor of income's negative correlation with impairment, because income rises over most of the range of ages in the sample and impairment also rises with age. Thus, adjusting for age actually increases income's coefficient compared to what it would be otherwise. Sex also has a significant direct effect in both samples, with women more impaired than men. Race and marriage have unstable effects, with each being significant in one sample but not the other. However, the signs of their coefficients remain consistent with the social background hypothesis, with whites and married people being less impaired.

REGRESSIONS OF CHANGE OVER TIME

The regressions of change over time serve three main purposes: they allow us to rule out pure selection as an alternative to the basic-needs hypothesis; they allow us to rule out a selection alternative to the resource-substitution hypothesis; and they provide evidence of selection effects that might contribute to the compounding of problems. We detail these and related points below. The over-time regressions all follow the same basic form, described in equation 1:

$$\Delta Y = b_0 + b_1 Y_{t1} + \sum_{i=2}^{k} b_i X_i + u_{\Delta Y} \tag{1}$$

In equation 1, b_i represents the effect of X_i on change in Y adjusting for its initial state at time one (Y_{t1}). The coefficient b_1 represents the instability of the outcome over the one-year period. Given that $\Delta Y = Y_{t2} - Y_{t1}$, equation 1 is algebraically equivalent to equation 2, where $1 - b_1$ represents the stability of the outcome:

$$Y_{t2} = b_0 + (1-b_1) Y_{t1} + \sum_{i=2}^{k} b_i X_i + u_{Y_{t2}} \tag{2}$$

Thus, the regression coefficients may be interpreted as effects on change net of initial state or as effects on outcome net of initial state. Only the coefficient of Y_{t1} differs in the two versions, representing instability in equation 1 and stability (1 minus instability) in equation 2.

Basic Needs

The previous cross-sectional results clearly show a negative association between income and health. The profile of the association seems consistent with the basic needs hypothesis. However, the association could result entirely from the impact of impairment on income. If physical impairment interferes with the ability to earn income, then some part of the cross-sectional association results from the social processes that move impaired persons downward in the income distribution. It is possible that *all* the cross-sectional association results from

selection and *none* of it from the health consequences of failing to meet basic needs. One way to rule out this alternative is to show that income predicts the rate and direction of subsequent change in impairment adjusting for initial level of impairment. The first column of Table 3 shows that regression, adjusting for social background and the initial levels of hardship and exercise.

Table 3 shows a significant negative effect of income on change in impairment. Thus, higher income predicts a lower rate of increase in impairment. Physical impairment tends to rise over time as people get older. The mean change in impairment is .032 on an initial mean level of .119, even though the Health Practices survey followed adults younger than 65 for only one year. The aggregate increase is significantly greater than zero at $p < .001$ ($t = 6.591$). The regression coefficient associated with the cube root of income is -.033, so a unit increase in the cube-root of income above its mean reduces the predicted change to zero. That level of income is about $42,600 in 1978 dollars, or $76,800 in 1990 dollars (the mean cube-root income is 2.494, which corresponds to 15.414 thousands; a cube-root income of 3.494 corresponds to 42.655 thousands; 1990 dollars are 1.8 times their 1978 equivalents). This theoretical income at which the rise of impairment vanishes is above the level of the code for the highest income category in the Health Practices data (30,000). Nevertheless, it is reasonable to infer that the expected rise of impairment tends to vanish at higher levels of income for adult Americans younger than 65.

The effect of economic hardship also tends to support the basic-needs hypothesis. Impairment rises more rapidly for those reporting earlier economic hardship. Of course, one indicator of economic hardship is not having enough money to pay for medical care, which could be a measure of health problems as much as of economic difficulties. The other indicator of economic hardship in the Health Practices data is any serious financial difficulty or problem over the five years preceding the initial interview. In a subsidiary regression analysis not shown in Table 3 we allowed forward stepwise regression to choose among the report of serious general financial difficulty, the report of difficulty paying for medical care, and the index averaging the two (probability for inclusion = .05, two-tailed). The procedure selected the index alone, indicating roughly equal effects of the two aspects of hardship on the change in impairment.

The direct effects of income and economic hardship on change in impairment support the basic-needs hypothesis. However, a progressive adjustment of income's effect on change in impairment does not show that hardship accounts for a large or significant fraction of income's total effect. The regression in the first column of Table 3 corresponds to the fully adjusted step 5 in Tables 1 and 2. A progressive adjustment of income's effect over time that are parallel to steps 1 through 4 in Tables 1 and 2 shows little change as adjustments are added. The regression coefficient is -.050 adjusting for initial impairment alone, -.044 adding adjustment for social background, -.038 adding adjustment for economic hardship, and -.034 adding adjustment for exercise. As in the cross-sectional analyses, the adjustments for social background and economic hardship reduced the size of income's coefficient the most. However, the changes are small compared to the coefficient's standard error of .011 — no larger than .545 standard errors. Thus, the changes with adjustment are all well within the 95% confidence limits of the original coefficient.

Resource Substitution

The cross-sectional results show that the strength of the negative association between impairment and income diminishes at higher levels of education. Selection forces could plausibly account for such a pattern if impairment reduces income less among those with higher education. For example, the incomes of manual laborers might be undermined by physical impairment far more than, say, the incomes of executives. Manual laborers tend to have high school education rather than college degrees. Thus, the stronger negative correlation between income and impairment could reflect a stronger effect of impairment on income at low levels of education, rather than a stronger effect of income on impairment. Two findings rule out this alternative interpretation. First, we find the same interaction between income and education as predictors of *change* in impairment: the effect of low income on change in impairment diminishes at higher levels of education, as shown in the right panel of Figure 4. Second, we find no indication of a significant interaction between education and impairment in *any* of the four over-time regressions, including the one predicting change in income. The p value for the education-by-impairment interaction term is .990 and .438 when the term is added to the equations predicting change in income and in economic hardship respectively.

Resource substitution also appears to occur with regard to economic hardship. Table 3 shows that education and income interact in their estimated effect on change in economic hardship. Higher education and income each predict relative declines in economic hardship, but less so in combination than implied by their separate effects. Put another way, low income tends to increase economic hardship, but less so the higher one's education.

Compounding Problems

Taken together, the regressions in Table 3 imply the compounding of problems over time. Physical impairment, income, and economic hardship significantly affect subsequent changes in each other. Impairment and exercise also significantly affect subsequent changes in each other. The signs of the coefficients all imply that changes tend to amplify rather than dampen. For example, impairment reduces income and exercise and increases economic hardship, all of which increases subsequent impairment. The result is a web of reinforcing effects — particularly among impairment, income, and economic hardship.

It is difficult to say exactly how much impact these self-reinforcing effects might have over a long period of time. Clearly, over a one-year period the indirect effect of one outcome on itself through the others cannot be large. To get a rough sense of the size of the one-year magnification we estimated a two-stage least-squares model of reciprocal effects between income and impairment. We used each outcome's initial value in 1979 as an instrument for its 1980 value in a standard reciprocal-effect model. The resulting coefficients imply that the reciprocal effect magnifies stochastic shocks by about .50 percent within the one-year period (that is, $[1 + b_{yx}\ b_{xy}]^{-1} = 1.005$). The annual rate seems small, but compounded over the 40 years of adulthood between ages 20 and 60 it would become a 22% enlargement of the initial effect ($1.005^{40} = 1.22$). For 10, 20, and 30

years of compounding the enlargement would be 5%, 10%, and 16% respectively.

Taken together, the over-time regressions indicate that physical and economic problems compound over time. The effect appears to be relatively small over a short period of time, but it could have substantial long-term consequences.

Status Lifestyle

The over-time regressions show that greater income predicts more positive changes in exercise. Average amounts of exercise declined significantly over the one-year period ($\Delta\bar{Y} = -.068$, $t = -3.237$, $p < .001$). The regression implies that the predicted decline in exercise is zero if income is $23,639 in 1978 dollars (or $42,552 in 1990 dollars) and all the other variables are at the sample mean. Exercise in turn significantly reduces the expected rise in impairment.

In regard to exercise, income appears to operate as a determinant of lifestyle rather than as an economic resource per se. Economic hardship has no significant effect on changes in exercise, but education has a positive effect as does income. Although exercise increases with social status, the over-time regressions give no indication that exercise helps people to rise in status, except to the extent that exercise forestalls impairment. Exercise has no significant net effect on changes in income or on changes in economic hardship controlling for impairment.

Social Background

Given the adjustment for baseline impairment, income, hardship, and exercise, social background has comparatively little effect on subsequent changes in the four outcomes. Not surprisingly, age has the biggest effects of the background variables, reflecting its role as an index of position in the life course. Older age accelerates the rise of impairment, the decline of exercise, and the decline of economic hardship, and slows the rise of income. Education also has effects on changes in all four outcomes, often modifying the effect of income. As noted in an earlier section, education and income operate as partial substitutes for each other in their effects on change in impairment and in hardship. Education also reduces the instability in income (or, alternatively, increases its stability), as indicated by a positive interaction with initial income in its effect on changes in income. Education also tends to slow the general decline in exercise over time. Table 3 shows only two other effects of background variables: physical impairment rises faster for women than for men, and household income rises faster for married people than for others.

Discussion

Our results confirm the sharp concentration of physical impairment near the bottom of the income distribution. By far the largest differences in physical impairment occur between the lowest and the next-to-lowest categories of income. Well over half the total decline in mean levels of impairment occurs between the bottom of the income distribution and the 6th percentile. Most of

the rest occurs between there and the 20th percentile. A number of results support the inference that the concentration comes in part from the deleterious effect of low income on health: a large fraction of the cross-sectional association survives adjustment for precursors that affect both income and impairment; a substantial fraction of the cross-sectional association vanishes with adjustment for economic hardship; lower income predicts steeper increases in physical impairment over time; and lower income predicts increased economic hardship and decreased frequency of exercise, both of which in turn predict sharper increases in physical impairment over time.

Basic Needs

Several findings suggest that the failure to meet basic physical needs accounts for part of low income's effect on impairment. These findings include the sharp concentration of impairment toward the bottom of the income distribution, the diminishing slope at progressively higher levels of income, the reduction of income's cross-sectional effect with adjustment for economic hardship, and the effects of income and hardship on change in impairment over time. Clearly, income reduces physical impairment in part because it helps people meet their basic physical needs for food, clothing, shelter, and care. Those needs appear to be met for most Americans at around the 20th percentile of household income — or about $20,000 a year in 1989 dollars. Beyond that critical point physical impairment declines relatively little at progressively higher levels of income.

Meeting basic needs partly explains the profile of impairment by level of income, but not entirely. Several other factors contribute to the pattern.

Resource Substitution

Education has two effects that combine to steepen the impairment slope toward the bottom of the income distribution and flatten it toward the top: education increases the level of household income and it decreases the effect of income on impairment. Income decreases impairment more for people with lower education — the very people concentrated toward the bottom of income's distribution. Income decreases impairment less for people with higher education — the very people concentrated toward the top of income's distribution. As a result, any given difference in income produces a greater effect on impairment toward the bottom of the socioeconomic scale than toward the top. Education reduces the need for money in the sense that it reduces the level of physical impairment, the rate of increase in physical impairment over time, the effect of income on the level of impairment, the effect of income on the rate of increase in impairment, and the effect of income on the rate of change in economic hardship. In part, the effect of greater income on impairment diminishes because education increases the typical level of household income while simultaneously diminishing the need for greater income.

Compounding Problems

The compounding of economic and health problems contributes to the increasing gradient of impairment at lower and lower levels of income. Income, economic hardship, impairment, and exercise form a web of mutually reinforcing effects that concentrate physical and economic problems together over time. Physical impairment slows or reverses the rise of income and the decline of economic hardship otherwise expected between the ages of 18 and 65. Lower income and greater economic hardship accelerate the rise of impairment. Similarly, impairment reduces the frequency of exercise, and lack of exercise speeds the rate of increase in impairment. The annual rate at which problems compound may be low — less than 1% by our rough estimate — but projected over decades of adulthood the effect could become substantial.

Status Lifestyle

Exercise's impact as an aspect of status lifestyle seems uncertain. On one hand the results clearly indicate that exercise forms part of a high-status lifestyle. Education and income both slow or reverse the decline of exercise over time. The effect appears to reflect lifestyle more than resources, because economic hardship has no effect on changes in exercise. Clearly, exercise predicts a slower rate of increase in impairment over time. On the other hand, adjusting for exercise in the cross-sectional analyses does not reduce the association between income and impairment. Furthermore, habitual exercise may come out of a high-status culture but it does not open doors to a high-status position. In the over-time analyses exercise does not improve income or reduce economic hardship, except indirectly by suppressing the rise of impairment. Outside its role in the web of compounding problems, exercise probably does not account for the concentration of impairment with low income.

Social Background

Shared social precursors account for 33.3% to 59.1% of the cross-sectional correlation between income and physical impairment. Education and sex play the biggest and clearest part, but race and marriage may contribute to the correlation too. Women have lower income and greater impairment than men, and women's impairment rises more rapidly over time. However, education operates as the single most important element of social background. Given the effects of education discussed in previous sections, education's background centrality comes as no surprise. The cross-sectional results attribute a remarkably large fraction of the association between income and impairment to fixed or highly stable social attributes — particularly education.

CONCLUSION

Why does income's incremental effect on physical impairment seem to diminish? Perhaps the question should be put the other way around. Why does the level of physical impairment rise ever more sharply toward the bottom of the income distribution? Looking at the results overall, a picture emerges. Education stabilizes by early adulthood, but its effects remain in force and

compound each other over decades, concentrating physical impairment and low income together. Throughout adulthood low education reduces income while sharpening the impact of low income on economic hardship and physical impairment. The effects compound over time, with impairment giving rise to greater hardship while further undermining income. Over the decades of adulthood each consequence of education accelerates changes in the others. By slowing the rise of income and speeding the rise of impairment, low education concentrates poverty and disability together.

If unmet needs and economic hardships produce physical impairments, then meeting the needs and reducing the hardships may sustain and improve physical functioning. In terms of impairment, the sharpest needs for additional income are limited to a minority of the population — no more than 20%. The amount of additional income needed appears to be in the range of $10,000 each. Suppose we could tax the rest of Americans to raise the money. How high would the tax be? There are four Americans above the critical level for every one below, so each one above would have to pay $2,500 in new taxes. The hard reality is that we in the upper 80% probably would not agree to pay so great a tax, even if we could be certain that it would eliminate the impairment of the needy. Education may provide the lever that can move this stone; it may lift people above the most serious needs by helping them to get higher income and to use their income more effectively.

References

Aiken, Leona S., and Stephen G. West. 1991. *Multiple Regression: Testing and Interpreting Interactions*. Sage.

Angell, Marcia. 1993. "Privilege and Health — What Is the Connection?" *The New England Journal of Medicine* 329:126-27.

Broom, Dorothy H. 1984. "The Social Distribution of Illness: Is Australia More Equal?" *Social Science and Medicine* 18:909-17.

Davis, James A. 1985. *The Logic of Causal Order*. Sage.

Diener, Ed, Ed Sandvik, Larry Seidlitz, and Marissa Diener. 1993. "The Relationship between Income and Subjective Well-Being: Relative or Absolute?" *Social Indicators Research* 28:195-223.

Epelbaum, Michael. 1990. "Sociomonetary Patterns and Specifications." *Social Science Research* 19:322-47.

Gold, Martha R., and Peter Franks. 1990. "The Social Origin of Cardiovascular Risk: An Investigation in a Rural Community." *International Journal of Health Services* 20:405-16.

Haghighatian, Mansour, and Linda Baer. 1990. "Cancer Mortality and Selected Social and Economic Factors, South Dakota: 1960 to 1980." *Wisconsin Sociologist* 28:13-18.

Hamilton, Lawrence C. 1992. *Regression With Graphics: A Second Course in Applied Statistics*. Brooks/Coles.

Hay, David Ian. 1988. "Socioeconomic Status and Health Status: A Study of Males in the Canada Health Survey." *Social Science and Medicine* 27:317-25.

Hayes, Diane, and Catherine E. Ross. 1986. "Body and Mind: The Effect of Exercise, Overweight, and Physical Health on Psychological Well-being." *Journal of Health and Social Behavior* 27:387-400.

Hirdes, John P., Stephen Brown, William F. Forbes, Debby S. Vigoda, Lawrence Crawford. 1986. "The Association between Self-Reported Income and Perceived Health Based on Ontario Longitudinal Study of Aging." *Canadian Journal of Aging* 5:189-204.

Hoffman, Mark S. 1991. *The World Almanac and Book of Facts*. Pharos.

House, James, Ronald Kessler, Regula Herzog, Richard P. Mero, Ann M. Kinney, and Martha J. Breslow. 1990. "Age, Socioeconomic Status, and Health." *The Milbank Quarterly* 68:383-411.

Johnson, Robert J., and Frederic D. Wolinsky. 1993. "The Structure of Health Status among Older Adults: Disease, Disability, Functional Limitation, and Perceived Health." *Journal of Health and Social Behavior* 34:105-21.

Maddox, George L., and Daniel O. Clark. 1992 "Trajectories of Functional Impairment in Later Life." *Journal of Health and Social Behavior* 33:114-25.

McDowell, Ian, and Claire Newell. 1987. *Measuring Health: A Guide to Rating Scales and Questionnaires*. Oxford University Press.

Mirowsky, John, and Catherine E. Ross. 1989. *Social Causes of Psychological Distress* Aldine de Gruyter.

Mosley, W. Henry, and Peter Cowley. 1991. "The Challenge of World Health." *Population Bulletin* 46:1-39.

Nagi, Saad Z. 1976. "An Epidemiology of Disability among Adults in the United States." *Milbank Memorial Fund Quarterly* 54:439-68.

O'Rourke, Diane, and Johnny Blair. 1983. "Improving Random Selection in Telephone Surveys." *Journal of Marketing Research* 20:428-32.

Pappas, Gregory, Susan Queen, Wilbur Hadden, and Gail Fisher. 1993. "The Increasing Disparity Between Socioeconomic Groups in the United States, 1960 and 1986." *The New England Journal of Medicine* 329:103-109.

Rogers, Richard G. 1992. "Living and Dying in the U.S.A.: Sociodemographic Determinants of Death among Blacks and Whites." *Demography* 29:287-304.

Ross, Catherine E, and Joan Huber. 1985. "Hardship and Depression." *Journal of Health and Social Behavior* 26:312-27.

Sen, Amartya. 1993. "The Economics of Life and Death." *Scientific American* 270:40-47.

U.S. Bureau of the Census. 1993. *Statistical Abstract of the United States 1993*. Government Printing Office.

Waksberg, Joseph. 1978. "Sampling Methods for Random Digit Dialing" *Journal of the American Statistical Association* 73:40-46.

Waldron, Ingrid, and Jerry A. Jacobs. 1988. "Effects of Labor Force Participation on Women's Health: New Evidence from a Longitudinal Study." *Journal of Occupational Medicine* 30:977-83.

Williams, David 1990 "Socioeconomic Differentials in Health: A Review and Redirection." *Social Psychology Quarterly* 53:81-99.

Contextual Influences on Young Men's Transition to First Marriage*

KIM M. LLOYD, *State University of New York at Albany*
SCOTT J. SOUTH, *State University of New York at Albany*

Abstract

Competing theories of marriage formation are evaluated by merging several contextual variables, primarily marriage market characteristics from the 1980 census, with men's marital histories observed between 1979 and 1984 in the National Longitudinal Survey of Youth. Discrete-time event history models reveal that, net of conventional individual-level predictors, a shortage of prospective partners in the local marriage market impedes white men's transition to first marriage. Women's aggregate economic independence, measured in terms of the proportion of females in the local marriage market who are employed and in terms of the size of average AFDC payments, also diminishes men's marriage propensities. Although earnings and home ownership facilitate men's marital transitions, racial differences in socioeconomic and marriage market characteristics account for relatively little of the substantial racial difference in marriage rates.

Patterns of family formation in the U.S. are undergoing substantial change. The latter half of this century has witnessed an overall retreat from marriage, with both the age at first marriage and the number of persons choosing to forgo marriage altogether increasing dramatically (Cherlin 1992). Although pervasive declines have occurred among all races, the retreat from marriage has been especially pronounced among African Americans. For example, in 1970 roughly the same proportion of black and white males had married by the age of 24 — 44% and 45%, respectively. However, by 1990, the proportion of black males who had married by age 24 had plummeted to 13%, compared to 22% of white males (U.S. Bureau of the Census 1991, 1993). Similar declines in marriage are found among white and African American females (Cherlin 1992).

Past research has not yet clearly identified whether increasing age at first marriage and increasing nonmarriage are caused by men's, women's, or both sexes' reluctance to form marital unions (South 1993). While the mutual nature of any marital decision is both obvious and straightforward, most discussions

* *This research was funded by a grant from the National Science Foundation (Grant No. SES-9121485). We thank Stewart Tolnay, Patty Glynn, and several anonymous reviewers for their helpful comments. Direct correspondence to Kim M. Lloyd or Scott J. South, Department of Sociology, State University of New York at Albany, Albany, NY 12222.*

 Social Forces, March 1996, 74(3):1097-1119

and empirical investigations of racial differentials and recent declines in marriage emphasize changes in *women's* perceptions and behaviors (e.g., Becker 1981; Espenshade 1985; Marini 1978; Schoen & Owens 1992). For example, ecological observations of marriage incidence (South & Lloyd 1992b) and prevalence (Preston & Richards 1975), individual-level longitudinal analyses of marital timing (Waite & Spitze 1981), and even contextual analyses of marital transitions (Lichter et al. 1992; McLaughlin, Lichter & Johnson 1993; Oropesa, Lichter & Anderson 1994) have, with few exceptions (Goldscheider & Waite 1986; Teachman, Polonko & Leigh 1987), focused exclusively on women's marital behaviors, thereby implicitly assuming that men's nuptial aspirations are less central to recent declines.

The present analysis makes three contributions to research in this area. First, by examining the determinants of men's marital transitions we provide a needed balance to previous studies that emphasize women's marital behavior. Second, we examine directly, and for the first time, the impact of marriage market characteristics and other contextual variables on men's transitions to first marriage while simultaneously controlling for conventional, individual-level predictors. This strategy allows us to evaluate competing models of the effects of the quantity and "quality" of females in men's marriage markets on their transition to first marriage. And third, we examine the extent to which racial differences in men's marital timing can be attributed to racial differences in demographic and socioeconomic characteristics, as well as to differences in marriage markets.

Factors Influencing Men's Transition to First Marriage

Marriage is one of the most important transitions that young men (and women) make as they enter adulthood (Goldscheider & Waite 1986). Marriage provides a clear indication of the passage from childhood to adulthood and is therefore a pivotal point in the life course. Theoretical explanations of marital transitions focus on both the marriage market conditions under which individuals search for a suitable spouse and individuals' own characteristics that make them either a desirable or an undesirable marriage partner.

MARRIAGE MARKET CHARACTERISTICS

Although few analyses have directly examined the impact of male marriage opportunities on men's marital behavior, theories of marriage timing are not completely silent on this issue. Theories of both marital search behavior and imbalanced sex ratios stress the numerical availability of potential partners but predict very different marital dynamics for men. Discussions of marital search behavior suggest that several aspects of the marriage market influence marriage propensities. Foremost among these is the numerical availability of potential partners. Local marriage markets that are characterized by discontinuity between the supply and demand for conjugal mates lead some prospective spouses either to remain unmarried or to delay the transition to marriage (Becker 1981; Oppenheimer 1988). It is important to note that theories of marital search behavior do not differentiate between males and females who face

unfavorable marriage market conditions. Therefore, a deficit in the quantity of potential wives should depress the first marriage transitions of males, just as a deficit of desirable potential husbands inhibits women's marriage (Lichter, LeClere & McLaughlin 1991; Lichter et al. 1992; South & Lloyd 1992b).

In contrast, theories of imbalanced sex ratios assert that when men are subjected to unfavorable marriage opportunities — that is, when there are more marriageable males than females in the local marriage market — men are *more* likely to marry (Guttentag & Secord 1983). Under these circumstances, women are in short supply and therefore men are motivated to commit to marriage in order to maintain a relationship with an opposite sex partner. Conversely, when men enjoy a favorable ratio of potential mates — that is, when there are more marriageable females than males in the local marriage market — men will be *less* likely to marry. Instead, it is argued, when there is an abundant supply of women, men utilize the increased number of female alternatives to avert or delay marriage (Fossett & Kiecolt 1990, 1993; Guttentag & Secord 1983). The relative surfeit of women available to men makes it unnecessary for men to commit to a single partner.

An interesting paradox exists when attempting to assess the relative merits of these competing theories. The overwhelming majority of research done on marital transitions, and virtually all studies of marriage market effects, have examined only women's marital timing. However, theories of marital search behavior and imbalanced sex ratios predict identical behaviors for *women* under similar marriage market conditions, and thus these studies cannot adjudicate between these competing theories. Both theories predict that women encountering numerous men in the local marriage market will have high marriage rates. Empirical investigations of women's marital behavior demonstrate that female marriage rates *are* positively related to the numerical supply of potential spouses, i.e., single men (Lichter et al. 1992; McLaughlin, Lichter & Johnson 1993; South & Lloyd 1992b). But these theoretical perspectives do predict a different relationship between mate availability and men's marriage propensities. Because there have been few empirical investigations of the impact of marriage opportunities and men's transition to marriage, however, it is not certain whether a surfeit of eligible women increases men's propensity to marry, as the marital search model would predict, or decreases the probability that men will form conjugal unions, as theories of imbalanced sex ratios assume.[1] This apparent conflict between competing theories justifies a focus on the influence of marriage market characteristics on men's marital transitions (Lichter et al. 1992).

Empirical investigations of *women's* marital timing demonstrate the importance of considering the "quality" of available mates, as well as their sheer quantity (Fossett & Kiecolt 1991; Lichter et al. 1992; South & Lloyd 1992b; Wilson 1987). Of particular importance for hastening *women's* entry into marriage is the employment and educational status of potential husbands. There are also reasons to anticipate that the characteristics of potential wives in the local marriage market will affect men's marital transitions, although contradictory expectations can be derived from the literature. On the one hand, South (1991) demonstrates that recent cohorts of unmarried men express a verbal preference for employed over unemployed wives. Mare and Winship (1991) find

that African American women who have good labor market prospects, as indicated by a high level of educational attainment and high weekly earnings, are more attractive to potential husbands. Thus, greater female labor force participation and higher levels of female educational attainment may increase *men's* gains from marriage, making employed and highly educated women more attractive marital prospects. This line of argument implies that male marriage propensities will be higher in marriage markets that contain a higher proportion of women who are employed or who have relatively high levels of completed schooling.

On the other hand, research focusing on *female* marriage transitions stresses the opposite dynamic by focusing on the role of women's economic independence as a deterrent to marriage formation. Theoretically, women will be less inclined to marry if they have attractive alternatives to the traditional role of wife and mother (Goldscheider & Waite 1986; Preston & Richards 1975; Teachman, Polonko & Leigh 1987; cf. Lichter et al. 1992). For example, women who are active in the labor force have greater financial security and therefore fewer economic incentives to marry. Additional support for this view is provided from theories of marital search behavior that suggest that favorable labor market status on the part of females should also depress male marriage rates because employment and education extend a woman's marital search process while simultaneously raising the "reservation quality" of her potential partner (Oppenheimer 1988). Presumably, the end result is that women who are employed or obtain high levels of education will be less likely to marry or will postpone their first marriage transitions.[2] If this is the case, areas with a high proportion of employed females, relative to areas with low levels of female labor force participation, should have lower rates of male marriage. Work by Fossett and Kiecolt (1993) support this position. They find an inverse relationship between black women's aggregate socioeconomic status and employment rate and the percentage of black men who are married across U.S. metropolitan areas. High levels of women's educational attainment should have a similar effect by increasing women's long-term labor market prospects, which thus reduces their anticipated gains from early marriage.

Three other characteristics of the local marriage market — the proportion of females enrolled in school full-time, average size of Aid to Families with Dependent Children (AFDC) payments, and population size — are hypothesized to diminish the probability that a man will marry. Women who are enrolled in school are less active in the marriage market primarily because marriage tends to threaten school completion. Past research consistently demonstrates that school enrollment inhibits marriage formation among both men and women (Goldscheider & Waite 1986; Teachman, Polonko & Leigh 1987). Second, discussions of declining marriage rates often identify the facilitating nature of welfare programs, asserting that they provide females with economic alternatives to male partnerships (Becker 1981; Murray 1984). Therefore, AFDC payments may be an important factor in establishing women's economic independence from traditional family formation. Consistent with this hypothesis is Fossett and Kiecolt's (1993) finding of an inverse effect of AFDC payments on black men's aggregate marriage prevalence across U.S. metropolitan areas. As does the proportion of women in the local marriage market who are currently

employed, who have a high level of education, or who are enrolled in school, AFDC payments serve as an explanatory variable that simultaneously reflects the probability that a male respondent will encounter an attractive and willing marriage partner and the degree to which women, in the aggregate, experience attractive alternatives to marriage. Finally, past studies suggest that urban community norms place less emphasis on traditional family values than do their rural counterparts, resulting in lower probabilities of marriage in densely populated areas (McLaughlin, Lichter & Johnson 1993; Teachman, Polonko & Leigh 1987).

MEN'S INDIVIDUAL-LEVEL CHARACTERISTICS

Prior research suggests that the most important individual-level characteristics facilitating men's transition to marriage are economic status, educational attainment, family background, and value orientation. Given men's traditional role as the family's primary financial provider, past analyses emphasize men's economic circumstances — specifically their labor market participation, job stability, and earnings — as pivotal determinants of marriage (Lerman 1989; Mare & Winship 1991; Oppenheimer, Kalmijn & Lew 1993; Wilson 1987). Economic characteristics are hypothesized to affect not only men's own current marital ambitions but also their attractiveness to potential wives (Oppenheimer, Kalmijn & Lew 1993; Wilson 1987). Men who establish a stable employment history and who earn high income are capable of providing and sustaining an independent household for themselves and their families. These attributes make men more attractive marriage prospects.

Other indicators of men's current and future wealth should also positively influence their marriage propensities. One signal of current financial stability is home ownership, while high levels of educational attainment may reflect future labor market potential and provide clues to long-term economic success (Oppenheimer 1988). Men who own their own homes and have high levels of educational attainment should therefore be more confident that they can provide for a family if they choose to marry. Furthermore, because of their increased socioeconomic status, these men should also find it easier to locate a willing marriage partner (Oppenheimer, Kalmijn & Lew 1993). School enrollment should negatively affect the probability of marriage for several reasons. Past research finds that, as with women, academic involvement inhibits marriage formation primarily because marriage tends to threaten school completion, which in turn may jeopardize a man's future career trajectory (Oppenheimer, Kalmijn & Lew 1993). The financial and time constraints of education probably operate to produce additional delays to marriage (Teachman, Polonko & Leigh 1987).

In addition to economic and educational characteristics, men's family background and value orientation also influence marital timing. Prior discussions of family formation suggest that individuals raised in disrupted family households or with single parents are more likely to delay first marriage (Avery, Goldscheider & Speare 1992), although contrary evidence can be found (Axinn & Thornton 1992; Goldscheider & Waite 1986). Finally, prior research also suggests that there are regional differences in family formation, with

southerners placing a greater emphasis on marriage and the family than do individuals residing in other regions of the country (Goldscheider & Waite 1986).

RACIAL DIFFERENCES IN MEN'S TRANSITION TO FIRST MARRIAGE

While numerous studies have explored levels of, and possible reasons for, the substantial difference in marriage rates between black and white females, the equally pronounced racial difference in men's marital timing has received comparatively little attention.[3] Explanations for the racial difference in female marriage rates have tended to emphasize the relative shortage of desirable black men to serve as husbands for black women (Lichter, LeClere & McLaughlin 1991; Lichter et al. 1992; South & Lloyd 1992b; Spanier & Glick 1980; Wilson 1987; cf. Schoen & Kluegel 1988). It seems unlikely, however, that racial differences in men's propensity to marry could be explained by racial differences in marriage market characteristics. Differences in mate availability between whites and African Americans imply that black men, relative to their white counterparts, enjoy a relative abundance of potential partners; if marital search theory is correct, this should lead to higher marriage rates for black than for white men.

It is possible, however, that racial differences in social and economic characteristics could account for some of the difference in the marriage rates of black and white men. Given the racial gap in such dimensions of socioeconomic status as income, employment, and home ownership, together with the tendency for these variables to increase men's propensity to marry, it is conceivable that some of the racial difference in men's marriage rates could be attributed to differences in these endowments. This view is consistent with Wilson's (1987; Wilson & Neckerman 1986) argument that African Americans' retreat from marriage is largely a consequence of black men's deteriorating labor market position. To examine this issue, we analyze the racial difference in men's timing of marriage both with and without controlling for an array of social and economic characteristics.

Data and Methods

The data for this study come from three sources: The National Longitudinal Survey of Youth (NLSY), the PUMS-D file of the 1980 census, and aggregate county-level data from the *U.S. County and City Data Book* (U.S. Bureau of the Census 1983). The NLSY is a national probability sample of 12,686 non-institutionalized individuals (of whom 6,403 are male), 14 to 22 years old at the time of their initial interview in January 1979. Respondents have been interviewed annually since then, although, for reasons discussed below, the present analysis will be confined to the 1979 through 1984 calendar years. The NLSY has many advantages over competing data sets for investigating young men's transition to marriage. First, the retention rate of the NLSY is high, with more than 90% of respondents remaining in the sample through 1984. Second, the NLSY contains extensive and detailed information on individual marital histories, labor market experiences, and socioeconomic characteristics. Third, the

survey covers the age categories in which most first marriages take place, i.e., 18 through 27. The primary advantage to using the NLSY to examine males' marital transitions, however, is the supplementary geocode files that identify the state and county in which each respondent resides at each annual interview, thereby allowing aggregated census data to be merged with each respondent's individual record.

This analysis restricts the NLSY sample in the following ways. First, following other recent analyses of marital transitions, we exclude individuals who married before age 18 (cf. Lichter et al. 1992; Teachman, Polonko & Leigh 1987). Marriages occurring to males below the age of 18 are quite rare and generally considered nonnormative. Moreover, data on many labor force and human capital characteristics are unavailable for males younger than age 18. Second, we focus on the marital experiences of non-Hispanic whites (hereafter referred to simply as whites) and African Americans. Hispanics display high rates of exogamy, marrying both non-Hispanic whites and African Americans at relatively high frequencies, thus making it difficult to delineate clearly their field of potential mates (Bean & Tienda 1987; Oropesa, Lichter & Anderson 1994). Third, in order to match geocode data from each annual interview, we include only respondents who are eligible for their first marriage between 1979 and 1984. Individuals who were already married before the initial 1979 interview are excluded from the analysis because we lack geocode information on the marriage market in which these marriages were initiated. Marriages that occurred after 1984 are also excluded because of the decreasing temporal accuracy of 1980 census data as measures of post-1980 marriage market conditions (Lichter et al. 1992). And fourth, because so few of the respondents in the early waves of the NLSY experience a second marriage, we focus solely on first marriage transitions. These selection criteria result in an effective sample of 3,907 young men — 2,712 non-Hispanic whites and 1,195 blacks.

DEFINING MEN'S MARRIAGE MARKETS

Most analyses exploring the effects of mate availability on family formation processes recognize that marriage markets are geographically circumscribed. Men (as well as women) search for, and select, marital partners from within local areas. Past research, however, is equivocal on what level of geography is the most appropriate, with various studies utilizing Metropolitan Statistical Areas (South & Lloyd 1992a; White 1981), urban areas (White 1979), counties (Fossett & Kiecolt 1990), and states (Schoen & Kluegel 1988; South & Lloyd 1992b). Following other recent analyses of marriage market effects (Lichter et al. 1992; McLaughlin, Lichter & Johnson 1993), we use data from the 1% PUMS-D file of the 1980 census (Tolbert & Killian 1987) to calculate mate availability measures across the 382 U.S. Labor Market Areas (LMAs). As others have noted (Lichter et al. 1992, South & Lloyd 1995), LMAs have several important advantages over rival operationalizations of marriage markets. First, LMAs are constructed around journey-to-work patterns and represent the spatial boundaries within which daily social action occurs. This quality makes LMAs a clearly superior representation of marriage markets than larger geographical areas such as states or nations. Second, unlike counties, LMAs can intersect state bound-

aries and, unlike Metropolitan Statistical Areas (MSAs), they encompass the entire U.S. population. Moreover, the microlevel PUMS-D file provides the necessary detail with which to compute appropriate measures of the quantity and quality of spousal options.

In addition to concerns about the proper level of geography used to define the local marriage market, issues of the appropriate measurement of prospective spouses must also be addressed. Clearly, not all individuals participate in the same marriage market. Two of the primary characteristics that segment marriage markets in the U.S. are age and race. Although significant differences in spousal ages may occur, they represent only a small minority of all marriages, with the overwhelming majority of people marrying individuals who are of roughly the same age. This is particularly true among younger cohorts of men marrying for the first time (Schoen & Weinick 1993). The modal pattern is for men to marry women approximately 2 years younger than themselves. And the vast bulk of marriages involve husbands and wives of the same race. Accordingly, when calculating the number of potential wives available to each male respondent, we require that "suitable" partners be single, of roughly the same age (see below), of the same race, and part of the noninstitutionalized population. Specifically, the ratio of potential spouses in each respondent's local marriage market is calculated as

$$RPS_i = \frac{\sum_{i-5}^{i+1} F_i}{\sum_{i-3}^{i+3} M_i} * 100 \qquad (1)$$

where RPS_i is the race-specific ratio of potential spouses for each male respondent age *i*, F_i is the number of unmarried females age *i* available to each respondent in his local marriage market, and M_i is the number of unmarried males age *i* in the respondent's local marriage market with whom he must "compete" for the females in the numerator. Because this article is concerned with the marital behaviors of men, we define the number of potential mates as the number of available females divided by the number of available males, multiplied by 100. This is simply the reciprocal of the traditional sex ratio. Thus, the relevant sex ratio for each male NLSY respondent is based on 7-year age groups in which men choose potential wives who are, on average, two years younger. For example, the relevant ratio of potential spouses for a 25-year-old white respondent is the number of single white females 20 to 26 years of age divided by the number of single white males 22 to 28 years of age. These ratios are calculated separately for each NLSY respondent across each of the six years covered in the analysis and are therefore sensitive to changes in respondents' ages as well as changes in their geographical location. Because these ratios are also based on broad seven-year age intervals, they are preferable to measures based on narrower age ranges (Fossett & Kiecolt 1991). And, while undercount of African American males in the 1980 census was not trivial, Fossett and Kiecolt (1991) demonstrate that such an underreporting will not distort spatial differentials or bias the observed effects of marriage market characteristics on processes of family formation.

We make two modifications to this sex ratio. First, so as not to treat very young females as eligible for marriage, the ratios are truncated slightly at the youngest ages. Specifically, the ratios of potential spouses for men age 18 and 19, respectively, are the number of females 16 to 19 divided by the number of males 18 to 21, and the number of females 16 to 20 divided by the number of males 18 to 22. Second, values of the ratio of potential spouses below the 5th and above the 95th percentile are recoded to those points in the age-race-specific distributions in order to reduce the impact of extreme observations, which most likely occur because of sampling error in smaller LMAs. In our sample, the "average" observed black marriage market, which is simultaneously delimited by LMA and age group, contains 274 unmarried black women and 215 unmarried black men. The corresponding counts for the typical white NLSY respondent in our sample are 689 unmarried white women and 690 unmarried white men. In the present analysis, 170 LMAs contain at least one black male NLSY respondent, while 332 LMAs contain at least one white NLSY respondent.

Two features of the research design may militate against finding significant effects of contextual explanatory variables for blacks. First, the comparatively small number of LMAs containing a black male respondent means that the full range of marriage market conditions experienced by the black population will not be captured in these data. If measures of mate availability vary by the likelihood that an LMA is represented in the analysis — e.g., if LMAs with small black populations (and thus no NLSY respondent) have unusual values on the sex ratio measures — then the variation in these variables will be truncated. Second, estimates of the values of the contextual variables may be less reliable for blacks than for whites because of the smaller sample sizes used to compute the former. Therefore, the observed contextual effects on black men's marital timing should be interpreted cautiously.

As noted earlier, it seems likely that, in addition to their relative supply, the social and demographic characteristics of potential wives might also influence young men's transition to first marriage. Particularly important are the numbers of women who are employed, have high educational attainments, and are enrolled in school. Accordingly, we compute from the PUMS-D file the percentage of unmarried women in each respondent's marriage market who are currently employed, have at least some college education, and are enrolled full-time in school. That is, these three indicators of the "quality" of potential wives are the percentages of women in the numerator of each respondents' sex ratio who have these characteristics. These time-varying covariates are measured in 1980 and are attached to each record by the respondent's residence at each annual interview in the same manner as the ratio of potential mates described above.

Two additional characteristics of the local marriage market are merged with the individual NLSY records. These are the population size of the LMA in which the respondent resides and the average value of state AFDC payments. Each is allowed to vary as the respondent changes residence from year to year. LMA population size is obtained from the PUMS-D file and included in the analysis because past research suggests that individuals who reside in heavily populated areas experience lower probabilities of marriage (McLaughlin, Lichter & Johnson 1993). Average AFDC payments are obtained from the *County and*

City Data Book (U.S. Bureau of the Census 1983). If AFDC payments provide women with alternatives to marriage, as the literature suggests, then men should find it harder to locate a desirable and willing marriage partner in geographical areas with higher average AFDC payments.

INDIVIDUAL-LEVEL COVARIATES

The remaining explanatory variables reflect economic, social, and demographic characteristics of the male respondents. Most of these variables also appear in prior analyses of men's marital timing (Goldscheider & Waite 1986; Mare & Winship 1991; Oppenheimer, Kalmijn & Lew 1993; Teachman, Polonko & Leigh 1987). Indicators of socioeconomic status are respondent's total annual income (in thousands of dollars), number of weeks worked in the year preceding the interview, whether the respondent owns his own home, and years of completed schooling. Full-time school enrollment and whether the respondent resided with both biological parents at age 14 are both measured by dummy variables. Finally, to examine possible regional differences in the propensity to marry, we include a dummy variable for southern residence. With the exception of respondent's childhood family structure, each of these explanatory variables is treated as a time-varying covariate, measured at each annual interview.

METHODS

The analysis employs a discrete-time event history approach (Allison 1984). A discrete-time, rather than a continuous-time, model is preferred largely for the ease with which time-varying covariates can be incorporated. We constructed a person-year file containing 17,094 records, based on the marital experiences of the 3,907 men. Each observation corresponds to a calendar year. Time-varying covariates are defined at the beginning of each annual interval. Only observations at risk of experiencing a first marriage are included in the analysis. The dependent variable is a binary variable indicating whether the respondent marries during each annual interval. Logistic regression models are used to assess the impact of marriage market characteristics and other explanatory variables on the probability of men's first marriage transitions (Yamaguchi 1991).

Results

Figure 1 presents life-table estimates of the cumulative percentage of respondents marrying by a given age, separately for white and black men. As anticipated, there are substantial differences in marriage propensities between white and African American men. Roughly 73% of white males are expected to marry by age 28, compared to only 48% of black males. These estimates are generally consistent with cross-sectional census data indicating sharp racial differences in men's marital status (U.S. Bureau of the Census 1985) and are also similar to previously observed racial differences in marital timing among young cohorts of females (Lichter et al. 1992). The racial difference in the transition to first marriage is further demonstrated by Table 1, which displays the means and

standard deviations, by race, for all variables included in the event-history analysis of men's transition to first marriage. These person-year observations demonstrate that, among those in a position to marry, white men are nearly three times as likely as African American men to marry during any given year.

Table 1 also demonstrates that black and white men search for prospective partners in very different marriage markets. As indicated by the difference in the mean ratio of potential spouses, black men enjoy a marriage market that contains a larger number of potential wives. The pool of eligible black women is, on average, 30% greater than the pool of eligible black men, while the ratio for whites approximates parity. White men, however, hold a relative advantage in the economic "quality" of single women in their local marriage market, measured here as the percentage of females who are employed or who have some postsecondary schooling. More of the women in white men's than in black men's marriage pool are employed and have at least some college education.

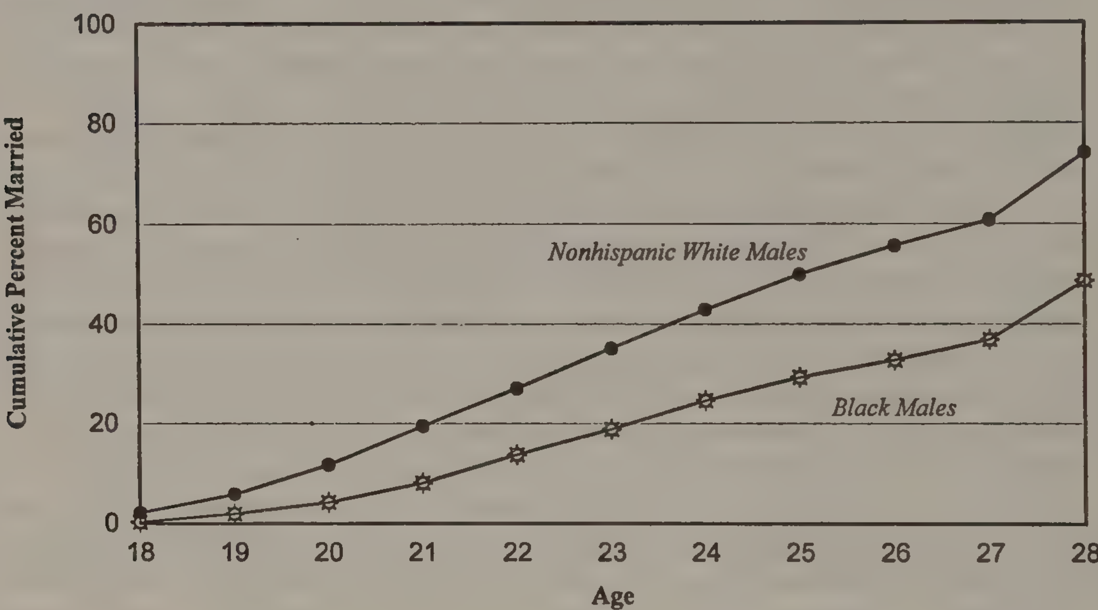
Consistent with prior research, white men also hold a decided advantage on the indicators of socioeconomic status (Mare & Winship 1991). Not surprisingly, white males tend to have higher annual incomes, have steadier employment, have more education, and have a greater probability of owning their own homes — although this latter endowment is relatively rare for both white and black respondents, most likely reflecting the relative youth of the sample. Such characteristics should reflect favorably on both the probability that men will choose to marry and, concomitantly, that they will find a willing partner.

Table 2 presents event-history models of white men's transition to first marriage. Model 1 is typical of previous individual-level analyses of men's marital entry that do not include measures of marriage market characteristics (e.g., Goldscheider & Waite 1986; Teachman, Polonko & Leigh 1987). Of the sociodemographic characteristics, school enrollment significantly impedes the transition to first marriage, while residence in the South enhances the likelihood of marriage. Growing up in an intact family is not significantly related to the timing of white men's first marriage. All four indicators of socioeconomic status significantly influence the annual probability of marriage: income, weeks worked, and home ownership significantly hasten the entry into marriage, while high levels of educational attainment delay it.

Model 2 in Table 2 adds the local marriage market characteristics and other contextual variables to the previous model. The addition of these variables significantly improves the model's ability to predict white men's transition to first marriage. Two of the marriage market characteristics — the ratio of potential spouses and the percentage of women who are employed — have significant effects. The relative number of eligible women in the local marriage market increases white men's annual probability of marriage. This effect is contrary to Guttentag and Secord's (1983) theory of imbalanced sex ratios but is consistent with a marital search model of mate selection. The larger the pool of eligible women, the greater the likelihood that men will marry.[4]

Despite recent evidence suggesting that men prefer wives who are active in the labor force (Mare & Winship 1991; South 1991), model 2 of Table 2 suggests that men are significantly less likely to marry when there is a relatively high percentage of single females in their local marriage market who are employed. As argued above, the financial independence that employment affords females

FIGURE 1: Life Tables Estimates of the Cumulative Percent Married by Age: Non-Hispanic White and Black Males Ages 18-27, 1979-1984



likely reduces women's incentive to marry and thus diminishes the probability that a young man will find a mate amenable to marriage (Goldscheider & Waite 1986; Preston & Richards 1975; Teachman, Polonko & Leigh 1987). The negative coefficient for average AFDC payments provides further support for this "independence effect." Men's likelihood of marriage is lower in counties with relatively generous welfare payments. However, neither the percentage of unmarried females with high education in the local marriage market nor the percentage enrolled in school significantly affects white men's marital timing. Finally, the effect of the remaining contextual variable, population size, is also significant and in the expected direction. The size of the LMA in which the respondent resides is inversely associated with the annual probability of marriage among white men.

Table 3 presents parallel analyses for the African American men in the sample. Again, model 1 is typical of previous studies that focus on individual-level determinants of the transition to marriage. As is the case for white men, school enrollment increases, while southern residence decreases, the likelihood of marriage. Socioeconomic characteristics appear to be important predictors of black men's marital timing. High income and home ownership both enhance significantly black men's probability of marriage. Unlike their white counterparts, however, educational attainment does not significantly affect black men's risk of marriage.

Model 2 in Table 3 adds the local marriage market characteristics and other contextual variables to the previous model of black men's transition to first marriage. Marriage market characteristics fail to emerge as significant determi-

TABLE 1: Mean and Standard Deviation of Variables Included in the Analysis of Men's Transition to First Marriage: Non-Hispanic Whites and Blacks, Ages 18-27, 1979-1984

Variable	Whites		Blacks	
	Mean	Std. Dev.	Mean	Std. Dev.
Marriage (0 = no; 1 = yes)	.08	.27	.03	.18
Social/demographic characteristics				
Age	20.86	2.13	20.80	2.11
In school (0 = no; 1 = yes)	.35	.48	.28	.45
Intact family (0 = no; 1 = yes)	.78	.41	.53	.50
South (0 = no; 1 = yes)	.28	.45	.54	.50
Socioeconomic characteristics				
Annual income (in thousands)	6.06	6.34	4.05	5.28
Weeks worked	30.01	19.58	23.42	19.97
Home ownership (0 = no; 1 = yes)	.04	.19	.02	.15
Education	12.32	1.91	11.68	1.73
Local marriage market				
Ratio of potential spouses	100.74	14.38	131.10	34.62
Percentage of single females employed	55.67	12.70	32.57	12.68
Percentage of single females with post-high school education	36.49	15.31	26.85	11.69
Percentage of single females in school	60.80	16.55	55.14	17.49
Other contextual variables				
LMA population size (in 100,000s)	22.21	29.53	28.22	34.71
Average AFDC payments	264.85	87.84	218.81	101.14
Number of person-years	11,696		5,398	

TABLE 2: Logistic Coefficients for Regression of Men's Transition to First Marriage on Selected Independent Variables: Non-Hispanic Whites, Ages 18-27, 1979-84

	Model 1		Model 2	
	Coef.	S.E.	Coef.	S.E.
Social/demographic characteristics				
Age[a]				
19	.20	.17	.23	.18
20	.56**	.16	.61**	.19
21	.77**	.16	.85**	.21
22	.87**	.17	1.03**	.29
23	.95**	.17	1.15**	.39
24	.77**	.19	.99*	.46
25	1.00**	.21	1.23*	.50
26	.78**	.26	1.03	.56
27	.61	.64	.90	.82
In school (0 = no; 1 = yes)	-.79**	.11	-.76**	.11
Intact family (0 = no; 1 = yes)	-.10	.08	-.10	.08
South (0 = no; 1 = yes)	.38**	.07	.02	.11
Socioeconomic characteristics				
Annual income	.03**	.00	.03**	.00
Weeks worked[b]	.40*	.20	.45*	.21
Home ownership (0 = no; 1 = yes)	1.48**	.11	1.44**	.11
Education	-.08**	.02	-.07**	.02
Local marriage market				
Ratio of potential spouses[b]			.53*	.24
Percentage of single females employed[b]			-1.33*	.56
Percentage of single females with post-high school education[b]			-.20	.53
Percentage of single females in school[b]			-.49	.73
Other contextual variables				
LMA population size			-.04*	.02
Average AFDC payments[b]			-.17**	.06
-2 Log likelihood	6,018.68		5,978.88	
Change in χ^2			39.81**	
Number of person-years	11,696		11,696	

[a] The omitted category is age 18.
[b] Coefficient and standard error multiplied by 100.

* $p \le .05$ ** $p \le .01$

TABLE 3: Logistic Coefficients for Regression of Men's Transition to First Marriage on Selected Independent Variables: Blacks, Ages 18-27, 1979-84.

	Model 1		Model 2	
	Coef.	S.E.	Coef.	S.E.
Social/demographic characteristics				
Age[a]				
19	1.09	.56	1.13*	.57
20	1.34*	.55	1.42*	.58
21	1.46**	.55	1.55**	.60
22	1.90**	.54	2.06**	.70
23	2.02**	.55	2.23**	.84
24	1.61**	.58	1.89*	.96
25	1.74**	.60	2.06*	1.01
26	1.33	.71	1.72	1.10
27	1.83	1.19	2.20	1.50
In school (0 = no; 1 = yes)	-.56*	.25	-.54*	.25
Intact family (0 = no; 1 = yes)	-.10	.16	-.13	.16
South (0 = no; 1 = yes)	.33*	.16	-.08	.25
Socioeconomic characteristics				
Annual income	.04**	.01	.05**	.01
Weeks worked[b]	.82	.44	.82	.44
Home ownership (0 = no; 1 = yes)	1.31**	.27	1.34**	.28
Education	.04	.05	.03	.05
Local marriage market				
Ratio of potential spouses[b]			.18	.26
Percentage of single females employed[b]			-.74	.95
Percentage of single females with post-high school education[b]			1.14	1.25
Percentage of single females in school[b]			.86	1.31
Other contextual variables				
LMA population size			.02	.03
Average AFDC payments[b]			-.33*	.15
-2 Log likelihood	1,483.92		1,474.73	
Change in χ^2			9.19	
Number of person-years	5,398		5,398	

[a] The omitted category is age 18.
[b] Coefficient and standard error multiplied by 100.

* $p \leq .05$ ** $p \leq .01$

nants of young black men's entry into marriage. As was the case among white men, however, the coefficient for the ratio of potential spouses is positive, and the coefficient for the percentage of single females in the local marriage market who are employed is negative. And, while these coefficients are not significantly different from zero, they are also not significantly different from the analogous coefficients for white men. One other contextual variable in model 2 — average AFDC payments — exhibits significant effects. As in the effects among white males, the size of average AFDC payments decreases black men's marriage probabilities, supporting a "female independence" model. LMA population size, in contrast, does not significantly influence the likelihood that black men will marry.

The analyses presented in Table 4 assess the degree to which racial (i.e., black versus non-Hispanic white) differences in young men's marital timing can be explained by racial differences in the independent variables. These equations are based on a pooled sample of person-year observations for both black and white respondents. The first equation in Table 4 includes as explanatory variables only the dummy variables for age and race. Exponentiating the coefficient for race in this equation implies that the risk of marriage among black males is only 41% ($e^{-.90}$) of the risk for white males. This difference is quite similar to the racial difference in young *women's* marriage probabilities observed by Lichter et al. (1992). The second equation in Table 4 adds the individual-level explanatory variables to the first equation. Controlling for these variables increases slightly the racial difference in the risk of marriage such that, net of these effects, black men are 37% ($e^{-.99}$) as likely as white men to marry in a given year. The third equation incorporates all the explanatory variables as predictors and, again, the racial difference in the risk of marriage tends to increase. In the face of all controls, the net risk of marriage for blacks is only 25% ($e^{-1.37}$) of the risk for whites. As a whole, then, racial differences in the explanatory variables tend to suppress, rather than to explain, racial differences in young men's marital timing.

Of course, the explanatory variables include both factors that might explain the racial difference in marriage probabilities and also factors that might suppress this difference. Variables that act to suppress the difference are education, school enrollment, average amount of AFDC payments, and percentage of females employed. Specifically, relative to white men, black men have lower levels of education and school enrollment and fewer employed women in their marriage markets, and they tend to live in areas with lower AFDC payments. In turn, each of these variables tends to reduce the likelihood of marriage. Hence, once these variables are controlled, the (standardized) racial difference in marriage probabilities grows larger. To a lesser extent, the racial difference in mate availability also "suppresses" the racial difference in marriage; black men have greater marriage opportunities and, in the sample as a whole, these opportunities tend, albeit modestly, to increase the likelihood of marriage.

However, three variables included in the model — income, weeks worked, and home ownership — could conceivably account for some of the racial difference in marriage, given that black men have lower means than white men on these variables and that all three variables hasten entry into marriage. But

TABLE 4: Logistic Coefficients for Regression of Men's Transition to First Marriage on Selected Independent Variables: Non-Hispanic Whites and Blacks, Ages 18-27, 1979-84

	Model 1		Model 2		Model 3	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Black	-.90**	.08	-.99**	.09	-1.37**	.16
Social/demographic characteristics						
Age[a]						
19	.48**	.16	.28	.16	.32	.17
20	.89**	.15	.61**	.15	.69**	.18
21	1.14**	.15	.80**	.15	.92**	.19
22	1.33**	.15	.97**	.16	1.18**	.25
23	1.49**	.15	1.06**	.16	1.34**	.34
24	1.39**	.16	.82**	.18	1.14**	.40
25	1.61**	.17	1.03**	.19	1.38**	.43
26	1.49**	.22	.77**	.24	1.16*	.48
27	1.25*	.54	.74	.56	1.17	.71
In school (0 = no; 1 = yes)			-.76**	.10	-.74**	.10
Intact family (0 = no; 1 = yes)			-.10	.07	-.11	.07
South (0 = no; 1 = yes)			.37**	.07	.01	.10
Socioeconomic characteristics						
Annual income			.03**	.00	.03**	.00
Weeks worked[b]			.50**	.18	.53**	.19
Home ownership (0 = no; 1 = yes)			1.46**	.10	1.42**	.10
Education			-.07**	.02	-.06**	.02
Local marriage market						
Ratio of potential spouses[b]					.32	.16
Percentage of single females employed[b]					-1.24**	.47
Percentage of single females with post-high school education[b]					-.16	.48
Percentage of single females in school[b]					-.14	.63
Other contextual variables						
LMA population size					-.03*	.01
Average AFDC payments[b]					-.18**	.06
-2 Log likelihood	995.84		7,526.77		7,485.41	
Change in χ^2			469.08**		41.36**	
Number of person-years	17,094		17,094		17,094	

[a] The omitted category is age 18.
[b] Coefficient and standard error multiplied by 100.

* $p \leq .05$ ** $p \leq .01$

racial differences in these three variables do little to explain the racial difference in marital entry. Adding only these three variables to the first equation of Table 4 reduces the coefficient for race from -.90 to -.76 (equation not shown), the latter figure implying that, even net of racial differences in income, weeks worked, and home ownership, black men experience a risk of marriage less than half that of whites ($.47 = e^{-.76}$). In short, it appears that very little of the racial difference in the propensity to marry can be attributed to racial differences in demographic, socioeconomic, or marriage market characteristics.

Conclusion

Patterns of U.S. family formation are changing as many individuals delay their first marriage or remain permanently single. Most theoretical and empirical investigations of marriage, however, emphasize only factors that affect women's decision to form conjugal unions. Consequently, the determinants of men's marital timing have received little attention. This neglect of factors influencing men's entry into marriage is especially problematic given evidence that young men express a greater reluctance to marry than do young women (South 1991). This analysis begins to bridge this gap in the literature by investigating the social context in which men make marital decisions. The primary focus has been on both the quantity and quality of the pool of eligible women in a man's local marriage market as determinants of men's transition to marriage. The results suggest that, at least for non-Hispanic white males, the probability of marriage increases as the number of single females increases. Such findings support a marital search model of mate selection that predicts that individuals will be more likely to marry when they encounter relatively numerous spousal options. A surplus in the quantity of females facilitates men's marital transitions by enhancing their assortative mating process. No support was found for theories of imbalanced sex ratios (e.g., Fossett & Kiecolt 1993; Guttentag & Secord 1983) that argue for the opposite dynamic, namely, that a surfeit of eligible women increases men's desire to avert marriage, thereby delaying such transitions.

The results also suggest that women's economic independence may serve as a deterrent to men's marriage formation. In general, marriage markets that are characterized by a high percentage of women who are employed and markets that have high average AFDC payments significantly decrease men's annual probability of marriage. Both these variables may represent factors that enable women to establish their economic independence from men and provide them with alternatives to the traditional roles of wife and mother. Within such marriage markets men may find it increasingly difficult to find a partner amenable to marriage. Women's labor market prospects may, however, have two offsetting effects. They may increase a woman's attractiveness as a potential wife (South 1991), while simultaneously reducing her economic incentives to marry. If these conflicting effects exist, it appears that the latter outweighs the former within the non-Hispanic white population. These effects may, however, counteract each other within the black population, thus leading to nonsignificant effects of female employment on the probability that black men will marry (Mare & Winship 1991). Further support for a female independence effect is

provided by the negative effect of AFDC payments on marriage probabilities. Both white and black men are less likely to marry in marriage markets that are characterized by high average AFDC payments.

This study also reveals significant effects of men's own economic and demographic characteristics on the timing of marriage. In general, economic resources such as income, employment stability, and home ownership accelerate the transition to first marriage, while school enrollment and high levels of educational attainment delay marital entry. The impact of men's economic resources suggests a partial reinterpretation of studies demonstrating a positive effect of the supply of eligible men in the local marriage market with desirable economic attributes on *women's* marital timing (Fossett & Kiecolt 1993; Lichter, LeClere & McLaughlin 1991; Lichter et al. 1992; South & Lloyd 1992b; Wilson 1987). These studies assume, either implicitly or explicitly, that the areal supply of men with valued economic resources increases women's propensity to marry because women prefer these characteristics in a husband, an assumption buttressed by survey data on women's mate selection preferences (South 1991). But this effect may also occur because economic resources increase *men's* propensity to marry, and thus the probability that a young woman will find in the local marriage market a man amenable to marriage. That is, in local marriage markets that contain a relatively high percentage of economically advantaged men, the marital preferences of both sexes reinforce each other, thereby increasing women's (and men's) marriage rates. This is in contrast to the situation described above, in which men's preference to marry employed women may be overridden by employed women's desire to delay marriage.

Finally, although several of the explanatory variables considered here exhibit significant effects on the timing of young men's transition to first marriage, very little of the pronounced racial difference in men's marriage propensities can be attributed to racial differences in socioeconomic or marriage market characteristics. In fact, standardizing for *all* the characteristics considered here increases, rather than diminishes, the racial difference in young men's likelihood of marriage. As with racial differences in female marriage rates (Bennett, Bloom & Craig 1989) and other dimensions of family structure (Morgan et al. 1993; Ruggles 1994), the racial difference in men's marriage propensities admits no simple explanation. While racial differences in economic factors account for some of the current difference and recent divergence (Mare & Winship 1991) in black and white men's marriage rates, they still leave most of the difference unexplained. Consequently, recent explanations have attempted to situate these broad racial differences within the unique culture and history of African Americans. Cherlin (1992), for example, suggests that blacks responded to a society-wide weakening of the institution of marriage and shifts in labor demand by drawing on the extended kinship networks that have historically been a significant component of the African American cultural repertoire. The reliance on extended kin provided resources that allowed blacks, more than whites, to delay or avert marriage in the face of societal changes inimical to the formation of nuclear families. However, this cultural explanation seems most relevant to explaining the retreat from marriage among black women; it is less clear how the availability of extended kinship ties would affect black men's marital propensities.

An alternative, but not incompatible, cultural explanation for racial differences in marriage emphasizes black men's attachment to peer groups whose values depreciate marriage. According to Anderson (1990), industrial restructuring has sharply curtailed the financial prospects and, consequently, marital prospects for inner-city black men. In turn, these limited marital prospects generate subcultural values that reject middle-class norms of marriage. While these values are initially developed and promulgated among lower-income groups, it seems reasonable to suggest that the strong attachment to peer groups among black men in general may transmit these values to black men having greater resources for marriage. Moreover, black men's low marriage rates, coupled with the tendency for friendship patterns to be homogamous with respect to marital status, imply that marriage will deteriorate peer group ties more among black than among white men. In short, we suggest that deteriorating economic prospects generate a rejection of marriage and that, because of strong peer group ties, these values diffuse throughout the population of young black men, even to those with sufficient resources for marriage. In contrast, peer group values of white men appear to be less averse to marriage (South 1993). Thus, among young men with similar resources for marriage, black men will be less motivated than white men to marry, as our results imply. If this explanation is correct, then studies of the impact of peer group attachment and the marital attitudes of significant others may prove useful in explaining the racial difference in the timing of young men's transition to marriage. In any event, it seems likely that a comprehensive explanation for the racial difference in young men's marital timing must look beyond demographic and economic differences to encompass social-psychological and cultural differences in how black and white men view the institution of marriage.

Notes

1. In one of the few investigations of this issue, Fossett and Kiecolt examine the effects of mate availability on marriage prevalence among African American men across nonmetropolitan Louisiana parishes (Fossett & Kiecolt 1990) and U.S. metropolitan areas (Fossett & Kiecolt 1993). In the former study, under the assumption of a linear relationship, they find positive associations between the sex ratio (men per 100 women) and the percentage of black men who are currently married, a finding consistent with Guttentag and Secord's (1983) theory of imbalanced sex ratios. Across metropolitan areas, however, they find that the sex ratio is inversely associated with black men's marriage prevalence, a finding more consistent with marital search theory. An important finding in both studies is the suggestive evidence that the association between mate availability and men's marriage prevalence is curvilinear, with the highest prevalence occurring where the numbers of men and women approach parity, and lower prevalences occurring in marriage markets where one sex outnumbers the other. Neither study, however, examines the effects of mate availability on the marital behavior of white men.

2. Prior studies have generated somewhat equivocal findings on the impact of female employment and education on women's marriage. Aggregate studies, using data on U.S. metropolitan areas or marriage markets, generally report *inverse* effects of women's labor force participation rate and mean years of school completed on the female marriage rate (Fossett & Kiecolt 1993; Freiden 1974; Lichter, LeClere & McLaughlin 1991; Preston & Richards 1975), while individual-level studies tend to find modest but *positive* effects of women's employment and educational attainment on their likelihood of marriage (Goldscheider & Waite 1986; Lichter et al. 1992; Waite & Spitze 1981). Whatever the reason for this incongruity, the aggregate analyses appear to be more relevant for the present analysis, and our hypothesis of inverse

effects of women's levels of employment and education on men's marital timing is consistent with this research.

3. The racial difference in marriage rates does not appear to be a consequence of higher nonmarital cohabitation rates among blacks than among whites. In fact, net of other factors, the likelihood of cohabitation is lower among African Americans (Bumpass & Sweet 1989).

4. Some interpretations of the sex ratio thesis imply a curvilinear relationship between the supply of potential wives and men's propensity to marry, with both a deficit and a surplus of mates inhibiting men's desire to marry and men's marital propensity peaking when the sex ratio approaches parity (Fossett & Kiecolt 1993). In additional analyses, we tested for this hypothesized nonlinearity with a polynomial response function, including as independent variables (in the equations in Tables 2 and 3) both the ratio of potential spouses in its original metric and its square. The coefficient for the square of the sex ratio was consistently nonsignificant, providing no support for the hypothesized curvilinear relationship.

References

Allison, Paul D. 1984. *Event History Analysis: Regression for Longitudinal Event Data.* Sage.

Anderson, Elijah. 1990. *Streetwise: Race, Class, and Change in an Urban Community.* University of Chicago Press.

Avery, Roger, Frances Goldscheider, and Alden Speare, Jr. 1992. "Feathered Nest/Gilded Cage: Parental Income and Leaving Home in the Transition to Adulthood." *Demography* 29:375-88.

Axinn, William G., and Arland Thornton. 1992. "The Influence of Parental Resources on the Timing of the Transition to Marriage." *Social Science Research* 21:61-85.

Bean, Frank D., and Marta Tienda. 1987. *The Hispanic Population of the United States.* Russell Sage Foundation.

Becker, Gary S. 1981. *A Treatise on the Family.* Harvard University Press.

Bennett, Neil G., David E. Bloom, and Patricia H. Craig. 1989. "The Divergence of Black and White Marriage Patterns." *American Journal of Sociology* 94:659-91.

Bumpass, Larry L. and James A. Sweet. 1989. "National Estimates of Cohabitation." *Demography* 26:615-25.

Cherlin, Andrew J. 1992. *Marriage, Divorce, Remarriage.* Harvard University Press.

Espenshade, Thomas J. 1985. "Marriage Trends in America: Estimates, Implications, and Causes." *Population and Development Review* 11:193-245.

Fossett, Mark A., and K. Jill Kiecolt. 1990. "Mate Availability, Family Formation and Family Structure among Black Americans in Nonmetropolitan Louisiana 1970-1980." *Rural Sociology* 55:305-27.

____. 1991. "A Methodological Review of the Sex Ratio: Alternatives for Comparative Research." *Journal of Marriage and the Family* 53:941-57.

____. 1993. "Mate Availability and Family Structure among African Americans in U.S. Metropolitan Areas." *Journal of Marriage and the Family* 55:288-302.

Freiden, Alan. 1974. "The U.S. Marriage Market." Pp. 352-71 in *Economics of the Family*, edited by Theodore W. Schulz. University of Chicago Press.

Goldscheider, Frances Kobrin, and Linda J. Waite. 1986. "Sex Differences in the Entry into Marriage." *American Journal of Sociology* 92:91-109.

Guttentag, Marcia, and Paul F. Secord. 1983. *Too Many Women? The Sex Ratio Question.* Sage.

Lerman, Robert I. 1989. "Employment Opportunities of Young Men and Family Formation." *American Economic Review* 79:62-66.

Lichter, Daniel T., Felicia B. LeClere, and Diane K. McLaughlin. 1991. "Local Marriage Markets and the Marital Behavior of Black and White Women." *American Journal of Sociology* 96:843-67.

Lichter, Daniel T., Diane K. McLaughlin, George Kephart, and David J. Landry. 1992. "Race and the Retreat from Marriage." *American Sociological Review* 57:781-99.

Mare, Robert D., and Christopher Winship. 1991. "Socioeconomic Change and the Decline of Marriage for Blacks and Whites." Pp. 175-202 in *The Urban Underclass*, edited by Christopher Jencks and Paul E. Peterson. Urban Institute.

Marini, Margaret Mooney. 1978. "The Transition to Adulthood: Sex Differences in Educational Attainment and Age at Marriage." *American Sociological Review* 43:483-507.

McLaughlin, Diane K., Daniel T. Lichter, and Gail M. Johnson. 1993. "Some Women Marry Young: Transitions to First-Marriage in Metropolitan and Nonmetropolitan Areas." *Journal of Marriage and the Family* 55:827-38.

Morgan, S. Philip, Antonio McDaniel, Andrew T. Miller, and Samuel H. Preston. 1993. "Racial Differences in Household and Family Structure at the Turn of the Century." *American Journal of Sociology* 98:798-828.

Murray, Charles. 1984. *Losing Ground.* Basic Books.

Oppenheimer, Valerie K. 1988. "A Theory of Marriage Timing." *American Journal of Sociology* 94:563-91.

Oppenheimer, Valerie K., Matthijs Kalmijn, and Vivian Lew. 1993. "Men's Transition to Work and Marriage Timing." Paper presented at the annual meeting of the Population Association of America, 1-3 April, Cincinnati, Ohio.

Oropesa, R.S., Daniel T. Lichter, and Robert N. Anderson. 1994. "Marriage Markets and the Paradox of Mexican American Nuptiality." *Journal of Marriage and the Family* 56:889-907.

Preston, Samuel H., and Alan T. Richards. 1975. "The Influence of Women's Work Opportunities on Marriage Rates." *Demography* 12:209-22.

Ruggles, Steven. 1994. "The Origins of African-American Family Structure." *American Sociological Review* 59:136-51.

Schoen, Robert, and James R. Kluegel. 1988. "The Widening Gap in Black and White Marriage Rates: The Impact of Population Composition and Differential Marriage Propensities." *American Sociological Review* 53:895-907.

Schoen, Robert, and D. Owens. 1992. "A Further Look at First Unions and First-Marriages." Pp. 109-17 in *The Changing American Family: Sociological and Demographic Perspectives*, edited by Scott J. South and Stewart E. Tolnay. Westview.

Schoen, Robert, and Robin M. Weinick. 1993. "Partner Choice in Marriages and Cohabitations." *Journal of Marriage and the Family* 55:408-14.

South, Scott J. 1991. "Sociodemographic Differentials in Mate Selection Preferences." *Journal of Marriage and the Family* 53:920-40.

____. 1993. "Racial and Ethnic Differences in the Desire to Marry." *Journal of Marriage and the Family* 55:357-70.

South, Scott J., and Kim M. Lloyd. 1992a. "Marriage Markets and Nonmarital Fertility in the United States." *Demography* 29:247-64.

____. 1992b. "Marriage Opportunities and Family Formation: Further Implications of Imbalanced Sex Ratios." *Journal of Marriage and the Family* 54:440-51.

____. 1995. "Spousal Alternatives and Marital Dissolution." *American Sociological Review* 60:21-35.

Spanier, Graham B., and Paul C. Glick. 1980. "Mate Selection Differentials between Whites and Blacks in the United States." *Social Forces* 58:707-25.

Teachman, Jay D., Karen A. Polonko, and Geoffrey K. Leigh. 1987. "Marital Timing: Race and Sex Comparisons." *Social Forces* 66:239-68.

Tolbert, Charles, and Molly Killian. 1987. *Labor Market Areas for the United States*. Department of Agriculture, Economic Research Service, Agricultural and Rural Economics Division.

U.S. Bureau of the Census. 1983. *County and City Data Book*. U.S. Government Printing Office.

____. 1985. "Marital Status and Living Arrangements: March 1984." *Current Population Reports* P-20, No. 399.

____. 1991. "Marital Status and Living Arrangements: March 1990." *Current Population Reports* P-20, No. 450.

____. 1993. "Marital Status and Living Arrangements: March 1992." *Current Population Reports* P-20, No. 468.

Waite, Linda J., and Glenna D. Spitze. 1981. "Young Women's Transition to Marriage." *Demography* 18:681-94.

White, Lynn K. 1979. "The Correlates of Urban Illegitimacy in the United States, 1960-1970." *Journal of Marriage and the Family* 41:715-26.

____. 1981. "A Note on Racial Differences in the Effects of Female Economic Opportunity on Marriage Rates." *Demography* 18:349-54.

Wilson, William Julius. 1987. *The Truly Disadvantaged*. University of Chicago Press.

Wilson, William Julius, and Kathryn Neckerman. 1986. "Poverty and Family Structure: The Widening Gap between Evidence and Public Policy Issues." Pp. 232-59 in *Fighting Poverty*, edited by Sheldon H. Danziger and Daniel H. Weinberg. Harvard University Press.

Yamaguchi, Kazuo. 1991. *Event History Analysis*. Sage.

SOCIAL FORCES

Forthcoming

AN INTERNATIONAL JOURNAL OF SOCIAL RESEARCH
ASSOCIATED WITH THE SOUTHERN SOCIOLOGICAL SOCIETY

VOLUME 74: NUMBER 4 JUNE 1996

Book Reviews

Social Change and the Experience of Unemployment.
Edited by Duncan Gallie, Catherine Marsh, and Carolyn Vogler. Oxford University Press, 1994. 373 pp. $59.00.

Reviewer: GARY P. GREEN, *University of Wisconsin, Madison*

Social science research on social change usually relies on one of two methodological approaches — analyses of large samples of a population, or case studies of communities, families, or individuals. Each has a set of strengths and weaknesses. This book provides an excellent model for bridging the gap between the two approaches. *Social Change and the Experience of Unemployment* is part of a series of volumes originating from the Social Change and Economic Life Initiative, a multidisciplinary project funded by the British Economic and Social Research Council that examines the consequences of the economic restructuring process on employer strategies, the worker's experience of unemployment, and the dynamics of household structures.

This volume is primarily concerned with the social causes and consequences of unemployment. Large-scale surveys were collected in six areas — Aberdeen, Kirkcaldy, Rochdale, Coventry, Northampton, and Swindon. All of these regions have experienced different patterns of economic and social restructuring. Other volumes in the series examine occupational change, employer strategies, gender segregation, the household, and trade unionism. Unlike many edited volumes, the various chapters in this book are integrated and hold together fairly well.

The authors of the introductory chapters do an excellent job discussing both the growing prevalence of unemployment in England and the different conditions in the six regions under study. Among the substantive topics examined in the book as a whole are the relationship of work attitudes to unemployment, the effects of the husband's unemployment on the wife's work, the effects of unemployment on psychological health, social networks and ties, marital dissolution, and welfare and collectivism. The analysis is timely given the recent interest in the U.S. in many of these topics.

U.S. researchers will value Gallie and Vogler's examination of work attitudes and unemployment. Their analysis is strengthened by the fact that they are able to examine the effects of work attitudes on unemployment over time. One of the major findings is that work attitudes among the employed and unemployed do not differ significantly and that work attitudes do not determine job status. The chapter by Davies, Elias, and Penn on the link between husband's unemployment and wife's labor force participation is also especially strong.

Although this is an impressive piece of work, I do not believe the researchers fully used the available data. Most of the community comparisons are lost in the analyses because the household data are aggregated. Obviously the aggregation was done for statistical reasons, but much of the context for economic and social

restructuring was lost as a result. In addition, because the project was broken down into several topics (e.g., skills, gender segregation, the household), the analyses of unemployment are separate from those of employer strategies, occupational change, and trade unionism. The researchers could have made more significant use of their unique data if, instead of combining the data in ways that resemble a national sample, they had conducted analyses demonstrating the linkage between the various substantive topics and the contextual processes of social change. Such analyses would have made a more valuable contribution to the literature on economic restructuring and unemployment.

Families in Troubled Times: Adapting to Change in Rural America.
By Rand D. Conger, Glen H. Elder Jr., and others. Aldine de Gruyter, 1994. 314 pp. Cloth, $46.95; paper, $23.95.

Reviewer: DORIS R. ENTWISLE, *Johns Hopkins University*

Like many books coming off the press these days, this one has multiple authors — 17 if I count correctly — with various authors assigned responsibility for individual chapters. Unlike many other such volumes, however, *Families in Troubled Times* has a clear focus and a common style — it is indeed a book rather than a compendium of chapters. Also, unlike many books that focus on a single research project, this one contains several chapters in which the research is placed in a larger context.

The focus of the research is on 451 intact families in rural Iowa, of whom many lost farms or farm-related livelihoods in the economic downturn in rural Iowa in the early 1980s. The data come from four members of each family: the father, the mother, an adolescent child, and the sibling closest in age to the adolescent were interviewed and observed.

In addition to a comprehensive description of the hardship-related experiences of the families, one coauthor (Lasley) traces the history of boom-and-bust cycles in agriculture and documents how the crisis in the 1980s led to severe out-migration and a general economic decline in the rural counties in which the study families resided. Lasley places the families in the larger social and economic context of rural America in the 1980s.

In describing how the families adapted to economic losses, Elder and associates describe the emergence of rural ghettos and relate how families made ends meet by cutting expenses and delaying purchases. They note trends toward off-farm employment and how children's work inside and outside the home helped counter adverse circumstances. They also describe the sense of hopelessness that overtook those who bore severe financial hardships. An intriguing chapter by Whitbeck and associates examines how the grandparent generation raised the parents, in particular how psychological weaknesses in the parents, which could be traced to their childhoods, undercut their ability to evoke needed social supports in the hard times of the 1980s.

LISREL-style models reveal, for example, that feelings of depression among wives or husbands were more prevalent in families with lower incomes, with incomes that were declining relative to prior years, with higher debt-to-asset ratios, or with unstable work patterns, although social and family support mitigated the adults' depressive symptoms. Nonetheless, economic stress exacted a heavy toll on marital quality and on parent-child relations. Economic pressures, besides having a

direct relationship to adolescents' adjustment, led to hostility between spouses that spilled over into explosive disciplinary practices with children. Parents' hostility even undercut warm and supportive feelings between siblings.

The collaborative research in *Families in Troubled Times* provides a nice counterpart to the shower of studies beginning to appear on poor families who live in urban settings. It differs from most of them by following traditions established in Elder's earlier work: invoking intraindividual psychological processes to explain behaviors prompted by economic loss, considering more than one parent in the same family, following a carefully defined cohort in a carefully defined social context over time, and, most of all, emphasizing the historical context surrounding these particular persons. In addition, it often pits one structural model against another so as to identify antecedent variables and their paths of influence as precisely as possible.

This book could be used in graduate or undergraduate classes, and it will be worth having over time.

From Redlining to Reinvestment: Community Responses to Urban Disinvestment. *Edited by Gregory D. Squires.* Temple University Press, 1992. 288 pp.

Reviewer: NANCY A. DENTON, *SUNY at Albany*

Two 20-year-old laws inspired what is wonderfully described in this book as "a quiet revolution." The 1975 Home Mortgage Disclosure Act (HMDA) requires the banking industry to release lending data by census tract, and the 1977 Community Reinvestment Act (CRA) defines the affirmative obligation of banks to meet their local community's credit needs. Thus the HMDA provides the data used to determine if the CRA obligation is met. While primarily aimed at researchers concerned with community development, housing, and neighborhoods, this book will also appeal to social scientists interested in social movements and in the effects of law on society. Seldom can the roots of a social movement be so clearly specified, and this movement has many and varied successes to its credit.

The first chapter, by Squires, and the last, by Calvin Bradford and Gale Cincotta, give superb theoretical, historical, and substantive context to the case studies that comprise the seven intervening chapters. Both should be read by everyone interested in the topic, and they would be useful class readings as well. The case studies focus on large cities, primarily in the industrial Northeast (Boston, Pittsburgh) and Midwest (Detroit, Chicago, Milwaukee), but they also include Atlanta and the state of California. Each is interesting and well written. The essence of the movement is the three-way partnership between government, banks, and the community to renew people and places. Both whites and minorities are victims of a system that discriminates against older communities, and they can unite, not in trying to change the system, but in trying to make it work for them. The communities vary in terms of success and failure, goals (banking services, banking location, single-family mortgages, multifamily mortgages, business loans), and strategies (e.g., the role of the media), so the reader gains a complete picture of how the HMDA and CRA can be used to aid neighborhoods.

For all the impressive positive accomplishments and empowerment reported in this work, the volume ends on a downbeat note as one learns how inadequately federal agencies enforce the CRA. Seldom do banks get bad CRA ratings (less than

15 percent of the time), and the regulatory agencies have obstructed the release of data on loans by race within each tract — data that might be used to identify areas on the brink of change for prompt intervention. Moreover, "not only do the regulators endorse the dual housing market, but they are willing to penalize those institutions that take aggressive steps to overcome conventional lending discrimination and act affirmatively on fair housing." As in the case of the 1968 Fair Housing Act, the failure of public policy is not in the laws but in their enforcement provisions and mechanisms. Communities are forced to do the work that a responsible government, one intent on fostering the principles of its own constitution, should do. Fortunately, this book shows that many communities are willing to do just that.

Chinatown: The Socioeconomic Potential of an Urban Enclave.
By Min Zhou. Temple University Press, 1992. 275 pp.

Reviewer: JAN C. LIN, *Amherst College*

The socioeconomic vitality exhibited by ethnic enclaves like Los Angeles's Koreatown and Miami's Little Havana has prompted reconsideration of traditional sociological assumptions regarding the urban ecology of immigrant settlement and the structural incorporation of immigrants into American society. Min Zhou's monograph on New York's Chinatown brings these issues into sharp focus as she confronts the assimilation model of the melting pot with the ethnic persistence paradigm of the enclave economy thesis.

Rather than viewing ethnicity as a liability that impedes socioeconomic and residential integration into American society, Zhou presents a more nuanced picture of ethnicity as a collective resource that is more enabling than constraining. Disadvantages of working within the ethnic enclave include low wages and exploitative working conditions; advantages include trust and loyalty, protection from societal prejudice and discrimination, viable employment despite limited proficiency in English, on-the-job training, potential for promotion, and opportunities for coethnic pooling of entrepreneurial capital. These tangible forms of social capital augment the workings of physical capital (money) in the enclave economy.

Zhou divides the enclave economy into a protected sector (oriented toward coethnic goods and service activities) and an export sector (oriented to the broader economy). Historically composed of laundry businesses, and replaced recently by the garment industry, the export sector is Chinatown's major source of income, which is recirculated through the protected sector via the Keynesian multiplier. Zhou overlooks the growing role of tourism as a source of export income for the enclave economy (via restaurants, curio shops, and jewelry stores).

Evidence for Zhou's study is primarily derived from field interviews and the public-use microdata sample (PUMS) of the 1980 U.S. census. Zhou uses a "returns on human capital" labor market perspective, following a methodology pioneered by Alejandro Portes in his examinations of Miami's Cuban and black enclave economies. Results of the analyses generally confirm that male enclave workers are able to gain positive returns on human capital (particularly college education). No specific differences are found, however, when comparing returns for participants inside or outside the enclave, or when comparing employees with entrepreneurs. Turning to gender, Zhou significantly finds neutral or negative returns on human

capital for women, leading to the conclusion that the broader positive functions of the enclave are partially derived from patriarchal exploitation.

Zhou suggests that the enclave "converts economic gains into residential ones," since a common sign of upward mobility among Chinese immigrants is the purchase of a home in New York's outer boroughs, commonly in "satellite Chinatowns" in Queens and Brooklyn. An analysis of data from the summary tape file of the 1980 census confirms this "ethnic resegregation" in the outer boroughs, which departs from traditional human ecological predictions that spatial assimilation will accompany socioeconomic mobility. There are also indications that satellite Chinatowns are absorbing some newer, poorer immigrants as well.

Finally, while Zhou correctly observes that overseas Chinese capital is a double-edged sword that augments the enclave economy while producing land-use conflict, she generally does not adequately contextualize her study in a world economy in which labor and capital are increasingly mobile. Her treatment of political ramifications is also weak, despite Chinatown's growing significance as a place of labor and protest action, a community-planning district for public officials, and a voting bloc in municipal elections. The sophisticated statistical evidence Zhou marshals in advancing ethnic enclave theory, however, makes her monograph one of the more important recent contributions to urban sociology, stratification theory, and race and ethnic studies.

The Population of Modern China.

Edited by Dudley L. Poston Jr. and David Yaukey. Plenum Press, 1992. 757 pp. Cloth, $79.50; paper, $44.50.

Reviewer: ELLEN EFRON PIMENTEL, *University of Illinois at Chicago*

The goals of this volume are laudable. Its editors aim to provide a compilation of present-day knowledge on the population of modern China suitable for students of world population and demography. The book is divided into eleven major sections covering such substantive areas as basic size and distribution, fertility trends and policies, international and internal migration and urbanization, mortality, marriage and family, and the process of demographic research in China. Each section contains two or three essays and includes a short introductory statement by the editors.

Given the explosion of recent research on Chinese population processes and policies, any book that promises to bring together major work on these topics sounds inviting. How well does this one fulfill its promise? Sociologist Nan Lin asserts in the foreword that a comprehensive appraisal of what is known about the Chinese population has so far been hindered by the lack of a single volume in which essential facts and analyses are made readily available. Perhaps this absence can be attributed to the heavy flow of data from China that continued into the late 1980s and early 1990s. In the context of the present volume, however, no such void exists, since the editors have largely reprinted essays using data from the early 1980s (with the exception of Poston's epilogue, which uses 1990 census data), covering much the same ground as Judith Banister in her excellent and comprehensive 1987 book, *China's Changing Population*. Indeed, Poston and Yaukey have reprinted four of Banister's 10 chapters, though her book is readily accessible. Given that the current volume was published late enough to include preliminary data from the 1990 census, it is curious that the editors did not delay publication in order to move

beyond Banister's contributions into more recent developments in Chinese demography.

More seriously, anyone interested in the goals outlined in the foreword — to examine how the Chinese population has been shaped by the interplays of political ideology and institutions, economy, government policies, sociocultural traditions, and ethnic divergences — will be unsatisfied. The breadth of the volume does not readily allow for such investigation. Though some essays do delve into these issues, the book is mostly a barrage of descriptive demographic data with no connecting theme or theory. While the value of having such a large quantity of data in one place is unquestionable, the lack of connecting themes makes the book seem somewhat disjointed and confusing.

The very attempt to provide information on such a wide variety of topics contributes to the sense of confusion. Each topic necessarily receives only superficial coverage. For example, except for the rather weak selection by Tien — an interview with Peng Peiyun — the section on fertility policy gives the reader little sense of the raging debates that have swirled around the implementation and consequences of the one-child policy, or of the resultant changes in policy throughout the 1980s and early 1990s. The demographic and social implications of differential regional implementation of the population policy are also ignored.

If one sticks to the very narrow goal of introducing students to the study of Chinese demography and to the basic findings of research in this field, then the book can be considered adequate to the task. Certainly this collection of essays includes many fine examples of basic data analysis from the 1982 census in particular, as well as other commendable studies from the 1980s. Especially effective is the combination of essays emphasizing empirical data on population trends and levels with essays fleshing out such background issues as ethnic identification and classification, problems of defining urbanization, and so forth. Greenhalgh's contribution, a behind-the-scenes look at how demographic research and collaborative work are carried out in China, is particularly admirable. Students of demography all too often are exposed only to the demographic facts and not to the processes by which those facts are brought to light or to the problems inherent in collecting demographic data.

The Rise and Fall of Catholic Religious Orders: A Social Movement Perspective.
By Patricia Wittberg. SUNY Press, 1994. 423 pp. $59.50.

Reviewer: DAVID YAMANE, *University of Wisconsin at Madison*

Although not without importance in the Catholic Church and society at large, the decline of Catholic religious orders has garnered little scholarly attention to date. Wittberg's book begins to fill this significant gap in the literature by providing an interpretation of religious orders as recurring movements of "virtuoso" spirituality. While the longest section of the book explores the rise of Catholic religious orders, the author's interest is clearly driven by their contemporary fall. Thus Wittberg opens by establishing the extent of the current crisis, offering some startling statistics: in the past 30 years the number of nuns and nonordained male religious in America has fallen over 50 percent; currently no new members are applying to 54

percent of all women's religious communities; and in 1990 the median age of nuns was 65.

To explain these statistics and the past 16 centuries of cyclical mobilization and demobilization, Wittberg draws on a variety of theoretical perspectives: resource mobilization theory to understand the material support needed in successful religious movements (chapters 4 and 5), frame alignment theory to understand the importance of ideology in successful mobilization (chapters 6-8), and work on commitment mechanisms in intentional communities (Kanter, Zablocki) to understand the internal dynamics of successful religious movements. She treats these bodies of scholarship as theoretical toolkits from which to draw the conceptual resources necessary to shed light on her evidence.

Wittberg does not, however, use her evidence to shed light on the theories. Consequently she passes over opportunities to make theoretical progress by specifying relationships between macrolevel processes (e.g., ideological frame alignment) and microlevel processes (e.g., communal commitment mechanisms), or between resources and ideology. In this regard *The Rise and Fall of Catholic Religious Orders* compares unfavorably to recent books on related topics, such as Finke and Stark's *The Churching of America* or Burns's *The Frontiers of Catholicism* (though it should be noted that Wittberg nowhere claims to be making a theoretical contribution).

Wittberg appropriates Weber's concept of "virtuoso" religiosity, which highlights the differential ability and willingness of individuals to pursue and attain sacred values. In the Catholic Church, religious orders are one of two institutionalized forms of virtuoso religiosity (diocesan priesthood being the other); thus new communities of religious are instances of successful movements of religious virtuosi. Wittberg identifies six varieties of religious virtuosi in the Catholic Church from the fourth through the twentieth century — desert monks, medieval monastics, Mendicants, Beguines, members of apostolic orders, and members of teaching congregations — each defined by the ideological frame expounded in promoting its innovative form of religious virtuosity.

In part 2 Wittberg takes the rise of these movements as her *explanandum* and seeks to identify the causal factors underlying their success. She draws attention to a variety of material and nonmaterial factors, both environmental and internal to the group, including demographic pressures (e.g., a shortage of marriageable men); support of ecclesiastical and secular authorities; an ideology resonant with the culture of the time and addressing a felt need or strain among some members of the population; opportunities provided by religious orders for education, financial security, and social status; social networks for evangelization; and basic material resources such as food, clothing, and shelter. While she invokes an array of factors and conditions suggested by social movement theory that are implicated in the successful establishment of religious orders, because there is no variance in her dependent variable (i.e., she does not consider any cases of unsuccessful movements to found religious communities), one cannot assess the relative importance of any of the factors.

Part 3 acts as a transition between the second and last sections. In it Wittberg suggests that the same factors that account for the rise of religious orders also determine their fall. As a prelude to her consideration of the contemporary period of decline, she highlights resource deprivation, frame dealignment, and the decay of communal commitment mechanisms as factors in the fall.

Part 4 is the most interesting. Wittberg's sustained treatment of developments during a single time period in a single national context — the U.S. in the postwar era — is more satisfying than her cursory coverage of the first 16 centuries of Catholic religious life throughout the Western world. She convincingly argues that contemporary religious orders in the U.S. find themselves facing extinction owing to a simultaneous delegitimation of the root ideology that had previously sustained them, a corresponding redefinition of everyday behavioral expectations toward individual preference and away from corporate commitment, and a loss of key resources such as ecclesiastical support and new recruits. Although Wittberg — herself a Sister of Charity — concludes by admitting a personal desire to see her own religious order avoid such a fate, if her analysis is correct, she and her fellow supporters of Catholic religious life have their work cut out for them.

This book does many things well. Among its strengths are its treating a previously neglected but consequential historical phenomenon, and its cross-national and historical breadth, which gives a sense of the diversity and importance of the form of religious expression Wittberg explores. Useful summary tables are interspersed throughout the text, and a brief glossary at the end provides a helpful resource for those not familiar with the language of religious orders.

A Weberian Theory of Human Society: Structure and Evolution.
By Walter L. Wallace. Rutgers University Press, 1994. 335 pp. $59.00.

Reviewer: DAVID NORMAN SMITH, *University of Kansas*

Walter Wallace has written a peculiar book, which appears under the auspices of the Rose Monograph Series of the American Sociological Association. It is neither "Weberian" nor a "theory of human society" in any meaningful sense, despite the promise of its title. Another, more significant eccentricity is Wallace's approach to his topic. His apparent goal is to offer a framework for the analysis of social evolution. Ordinarily a project of this kind entails immersion in ethnology and history, in the manner of Sanderson or Krader. Yet Wallace omits almost every matter of fact, offering instead a vague formal model of "social evolution" — a frankly deductive model that he grounds almost entirely in an idiosyncratic reading of Max Weber. This is a peculiar way to approach history, and an equally peculiar way to approach Max Weber.

Wallace opens by cheerfully confessing that his turn to Weber is prompted in part by a wish to be "relevant." Since Weber is almost universally acclaimed, there is, he reasons, a "prima facie assurance that a Weber-inspired theory is likely to address issues of continuing importance." He quickly concedes, though, that his "integration" of Weber's perspectives "is (of course) not to be found in Weber's published work itself." "I have," he writes, "more than half-invented this integration."

Wallace's move is problematic. Since he works without historical data, he is dependent on Weber for insight into history. Unfortunately, however, his Weber is more than half invented and not a good source of data on the past. The problem is not that Wallace is unable to read or grasp Weber. On the contrary, when Weber says what he wants to hear, Wallace is all ears. But Wallace's intellectual agenda is so all-consuming for him, and so exclusive of other sensibilities, that he seems tone-deaf when Weber shifts into another key. The result is that in many places (e.g., in

his account of the Protestant ethic) Wallace is far from the mark in his exposition of Weber's views. He is so eager to map the "phylogenetically given behavioral structure" of social conduct, for example, that he manages to find little more in Weber's anatomy of asceticism than the implication that the "raw materials for the entrepreneur's spirit of capitalism were already present in the human genetic endowment."

Wallace's exegesis of Weber is so eccentric that his technical grasp of Weber's texts is almost irrelevant. It is, after all, possible to be wrong about Max Weber but right about the world. Yet here, too, Wallace falls short.

To "explain" change, for example, Wallace invokes the category of charisma as an explanatory master key. But his notion of charisma is so vague that the term amounts to little more than a slogan, invoked whenever a "punctuation" in history is observed (or predicted). Wallace is thus reduced to empty generalities, affirming, for example, that his model shows "the possibility of an irregularly repeated process that predicts that *any* established stage of human societal culture is likely, sooner or later, to succumb to outbursts of high-intensity charisma . . . that . . . may then force a relatively sudden transition to a new and different stage — or back to a previous one."

Still, Wallace's outlook is so sunny that his book, for all its faults, has a certain charm. It is hard to resist the sheer optimism of an author who can find hope in the prospect of "a worldwide psychical contagion of the charisma of peace, a contagion that can provide a new and radically different punctuation in the future course of history." Yet hope is a poor substitute for historical vision. "Punctuations" do not spring from science, like Athena from the forehead of Zeus. To prepare seriously for the future, one must wrestle with history. Walter Wallace is right, at least, that Max Weber is a likely source of insight on this level.

The Semiotic Self.
By Norbert Wiley. University of Chicago Press, 1994. 263 pp. Cloth, $39.95; paper, $19.95.

Reviewer: ALAN SICA, *Pennsylvania State University*

Norbert Wiley has concluded a notable career as an American (*sic*) social theorist (Donald Levine's jacket blurb correctly notes a Whitmanesque quality to the book) by producing an unusual and challenging assessment of the interior, reflexively monitored monologue/dialogue known in simplified form to most theorists as "the self." He does this with multiple goals, from the most minutely micro to the global, and the unusual character of his project reflects a theoretical sweep considerably broader than many others within the so-called social-psychological realm. He probes around within the psyche while at the same time addressing larger political questions, speaking a language in elaborated form originally invented by his two principal inspirations from classical theorizing, Peirce and Mead, neither of whose conceptual innovations stayed carefully within the properly "psychic." That Wiley's "neo-pragmatic" posture invites his readers to reappropriate them both, particularly Peirce's theories of language and being — which accounts for the book's title being changed midstream from a Meadian "The Reflexive Self" to the published version, a change obviously inspired by Peirce — is but one of its virtues.

Wiley points out that Mead probably never read Peirce, and whereas Mead still enjoys an official coterie of theoretical followers under the banner of symbolic interactionism, Peirce is without many champions. Explaining why is not feasible within such brief compass, aside from noting that no matter how difficult, underdeveloped, and imprecise Mead's theorizing about the self/society mix may be, it bespeaks lucidity itself when compared with Peirce's thoughts on similar matters. For Wiley to sail under Peirce's semiotic flag is therefore risky: to the extent that his intended audience does not or will not embrace the latter's basic semiotic theory, Wiley's attempted merger of the two classical titans is jeopardized. Peirce's notions are thorny and represent the height of conceptual and terminological idiosyncrasy (e.g., "the fact that every thought is a sign, taken in conjunction with the fact that life is a train of thought, proves that man is a sign"; "the entire universe is perfused with signs, if it is not composed exclusively of signs"). When his key terms — firstness/icon, secondness/index, thirdness/symbol, the representamen, the interpretant, and orgies of trichotomization — are seriously evaluated for their contemporary utility as semiotic guides, particularly when compared with the reigning models (Saussure, Ogden and Richards, Barthes, Hjelmslev), it comes as little surprise that his lead has not been followed in analyzing consciousness. Furthermore, that Wiley cites a number of standard sources from Mead in his bibliography but not a single one from Peirce is puzzling. He seems to have relied almost entirely on Vincent Colapietro's *Peirce's Approach to the Self* (1989), a 125-page monograph that, while respected by pragmatist philosophers, can hardly be taken as synonymous with Peirce's original work, especially given that the book's avowed goal is to offer a new theory of the "democratic" self on the back of a merger between symbolic interaction and semiotics.

Wiley is nothing if not courageous: "I will argue that this [Peirce-Mead] synthesis has the strongest elective affinity with the times and is the best choice for the contemporary American situation. This argument will have three parts: that democracy works best if it has a theory of the self; that this theory should explain equality, freedom, and the inviolability of the self, and the Peirce-Mead synthesis is especially useful at this particular time in American history." To what extent Wiley's unique message will influence other theorists (or the political culture of America) probably depends as much on the ready understanding of his theory as on the "truth" it may contain. When he writes, "Humans are a triad of triads, and, in addition, the three triads merge into one. . . . I usually refer to them, in dialogical short-hand, as the I, you, and me, though the more precise names are I-present-sign, you-future-interpretant, and me-past-object," or when he explains, "All signs are semiotic triads, consisting of 'sign,' interpretant, and object, but humans are triads in a unique way. They are the signs behind the signs, or . . . bi-leveled signs," a wily (*sic*) blending of the ordinary with the insightful, not unlike Peirce's own conjectures, is clearly on display. And like most efforts at innovation, some time may pass before all its virtues can be known.

Old-Age Security in Comparative Perspective.
By John B. Williamson and Fred C. Pampel. Oxford University Press, 1993. 316 pp. $49.95.

Reviewer: JOHN A. HALL, *McGill University*

Specialized books on seemingly abstruse aspects of social policy rarely gain a general readership. This is often a great pity, because examination of such detail is what allows for high-powered understanding of how states and societies interact. It would be positively shameful for sociologists to ignore this marvelous book, which is a model for practitioners of sociology in almost every way. Substantively, the work is pioneering in that its authors consider within a single compass both advanced capitalist countries and developing states: separate chapters are thus devoted to Germany, the United Kingdom, Sweden, the U.S., Brazil, India, and Nigeria. It is equally striking methodologically: Williamson and Pampel place their essays on comparative historical sociology within the larger perspective of cross-national quantitative analysis, a method they have used in previous work. More generally, the book is exceptionally thorough and cognitively powerful, particularly in the way the authors seek to use data to assess theory.

Analysis begins, and ends, with the evaluation of five general theories of social-policy advance. Considered are the view that the requirements of industrialism dictate policy advance; the theory of social democracy, probably still dominant in the literature, which privileges pressure from below; neo-Marxist theorizings of rational states seeking to preserve capitalism from above; neopluralist views; and the recently influential state-centered analysis. One clear finding is that the social-democratic perspective explains very little. For one thing, it sheds small light on the surprisingly well-developed system of old-age provision in the U.S.; for another, the expansion of provision for old age in Germany has resulted for the most part from pressure from middle-class rather than working-class constituencies. Further, the only reason the perspective makes sense of the Swedish case is that the state encourages organized pressure from trade unions, which the authors suggest may be true of corporatist societies more generally. The authors also find that the theories in question, which variously do help explain policy in advanced capitalist countries, are applicable to the developing world. Just as ethnic division has an impact on policy provision in the U.S., so too does it affect policy considerations in Nigeria and India. At a general level, the neo-Marxist and industrialism-based perspectives best explain the policy decisions of the late developers.

Williamson and Pampel have written a superlatively synthetic book, making use of theory to explain reality, and have provided an absolute model of lucidity and rigorous argument. Still, they concentrate perhaps too exclusively on public provision; any final account of old-age security would also need to examine the ways that tax systems aid private provision for old age — most notably in Japan, analysis of whose idiosyncrasies is sadly missed. Further, the authors accept too readily standard classifications of corporatism, having not absorbed the reasoning and findings of Michael Smith's important *Power, Norms and Inflation* (1992). One wonders, finally, if more attention to purchasing-power parities would have changed the analysis, there being well-known problems with uncritical conversion into the dollar. But these are only minor criticisms of a major achievement.

Take Note

Blieszer, Rosemary, and Victoria Hilkevitch Bedford (eds.). *Handbook of Aging and the Family. Greenwood Press, 1995. 536 pp. $95.00.* Traversing the theoretical and methodological, this multidisciplinary reference examines family relationships and family life. With a focus on families with aging members, boundaries of families are broadly defined. Family life beyond the rearing of children, among members of different households, and among persons who no longer are connected through blood or marriage are considered.

Chen, Lincoln C., Arthur Kleinman, and Norma C. Ware. *Health and Social Change in International Perspective. Harvard University Press, 1994. 508 pp. Paper, $15.00.* The Health Transition Program conducted at the Harvard Center for Population and Development Studies from 1987 to 1990 generated this collection of papers focusing on the social dimensions of health change. The dialogue is interdisciplinary in nature and is oriented towards developing countries. The selections describe patterns of change in health, conceptual and methodological issues, the social dimensions of mortality and illness, and means of promoting health.

Ermisch, John, and Naohiro Ogawa (eds.). *The Family, the Market, and the State in Aging Societies. Oxford University Press, 1994. 239 pp. $42.00.* This combination of economic and demographic analysis explores the consequences of change in family formation and dissolution over the past twenty years. Individual chapters consider the U.S., Europe, and Japan, and are divided into two sections. The first considers market and state influences on family formation and the second concentrates on intergenerational transfers.

Thronton, Arland, and Hui-Sheng Lin. *Social Change and the Family in Taiwan. University of Chicago Press, 1995. 466 pp. $29.95.* The implications of rapid social and economic change for family change are considered. The patterns and causes of change addressed include education, work, income transfers, leisure time, marriage, and living arrangements. Perspectives are interdisciplinary and analyses are organized broadly around two conceptual frameworks — the modes of social organization framework and the life-course perspective.

Erratum

In Denise Kandel's article, "The Impact of Drug Use on Earnings: A Life-Span Perspective," published in the September 1995 issue (74:1), the last sentence of the second paragraph on page 261 should read as follows: "Finally, (c) drug users experience less growth in their wages than nonusers from the late twenties to the midthirties."

Referees

We record our gratitude to our colleagues who assisted us from July 1, 1994 through June 30, 1995. Their reviews of several hundred manuscripts provided a valuable service to the authors, to the journal, and to the profession. The list does not include people who were Editorial Board members during that period.

Acock, Alan C.
Adams, Julia P.
Alba, Richard D.
Aldous, Joan
Alexander, Karl L.
Alexander, Victoria D.
Alwin, Duane F.
Amato, Paul R.
Aneshensel, Carol
Angel, Ronald J.
Anheier, Helmut
Appold, Stephen J.
Ardelt, Monika
Arjomand, Said
Armer, J. Michael
Astone, Nan Marie
Baker, David
Baker, Wayne E.
Baum, Joel A.C.
Bayer, Alan E.
Beck, E.M.
Beisel, Nicola
Bellas, Marcia
Benford, Robert D.
Best, Joel
Bian, Yanjie
Bianchi, Suzanne
Bielby, Denise D.
Bielby, William T.
Bielefeld, Wolfgang
Biggart, Nicole W.
Birnbaum, Norman
Bokemeier, Janet L.
Boli, John
Booth, Alan
Bowden, Gary
Boyd, Robert L.
Bradshaw, York W.
Breedlove, William L.
Brint, Steven G.
Brody, Charles J.
Broman, Clifford L.
Budros, Art
Burns, Gene
Burr, Jeffrey A.
Bursik, Robert J., Jr.
Burstein, Paul
Calasanti, Toni M.
Camic, Charles
Campbell, John L.
Carley, Kathleen
Cernkovich, Stephen A.
Chafetz, Janet Saltzman
Chaves, Mark
Cherlin, Andrew J.
Christiano, Kevin J.
Clemens, Elizabeth S.
Cochran, John K.
Collins, Randall
Cooney, Teresa M.
Corse, Sarah M.
Corzine, Jay
Coxon, Tony
Cramer, James C.
Crenshaw, Edward
Danigelis, Nicholas L.
Davis, Deborah
Davis, Gerald F.
Davis, James A.
Davis-Blake, Alison
DeBord, Larry W.
Delacroix, Jacques J.
Demo, David H.
Denton, Nancy A.
Devine, Joel
Dillon, Michele
DiMaggio, Paul J.
DiPrete, Thomas A.
Dowd, James J.
Dubin, Steven C.
Eberstein, Isaac W.
Eggebeen, David J.
Ellison, Christopher G.
Epstein, Steven G.
Falk, William W.
Farley, John E.
Farley, Reynolds
Felson, Richard B.
Fiala, Robert A.
Finke, Roger K.
Firebaugh, Glenn
Fligstein, Neil
Ford, Rebecca
Form, William
Frey, R. Scott
Fugita, Stephen S.
Galaskiewicz, Joseph
Gamoran, Adam
Goe, W. Richard
Gold, Thomas B.
Goldstone, Jack A.
Goode, Erich
Grandjean, Burke D.
Granovetter, Mark S.
Grant, Don Sherman, II
Grasmick, Harold G.
Green, Gary P.
Greenstein, Theodore
Griffin, Larry J.
Guest, Avery M.
Hagstrom, Warren O.

Hall, John R.
Hall, Thomas D.
Hamilton, Gary G.
Hansell, Stephen E.
Hargens, Lowell L.
Hargis, Peggy
Hartman, John
Hasenfeld, Yeheskel
Hawkins, Darnell F.
Hayward, Mark
Heaton, Tim B.
Hegtvedt, Karen A.
Henretta, John C.
Hicks, Barbara E.
Himes, Christine
Hondagneu-Sotelo, Pierrette
Hooks, Gregory
Hout, Michael
Hughes, Michael
Hummer, Robert A.
Hurlbert, Jeanne S.
Iannaccone, Laurence
Idler, Ellen L.
Isaac, Larry W.
Ishii-Kuntz, Masako
Jackman, Mary R.
Jackson, Pamela Braboy
Jacobs, Jerry A.
Jaffee, David D.
Jasper, James M.
Jenness, Valerie
Jiobu, Robert M.
Johnson, Cathryn
Kamo, Yoshinori
Kandel, Denise B.
Kane, Emily W.
Kaufman, Robert L.
Kennedy, Michael D.
Kessler, Ronald C.
Kiecolt, K. Jill
Kilbourne, Barbara Stanek
Kluegel, James R.
Knoke, David
Kohn, Melvin L.
Krecker, Margaret L.
Krivo, Lauren J.
LaGory, Mark E.
Land, Kenneth C.
Lang, Gladys Engel
Lechner, Frank J.
Lehrman, William G.
Liao, Tim Futing
Lichter, Daniel T.
Light, Ivan
Lin, Nan
Lofland, John F.
Lofland, Lyn H.
Logan, John R.
Long, Gary L.
Long, J. Scott
Longshore, Douglas
Lorence, Jon
Loscocco, Karyn A.
Lovaglia, Michael J.
Lye, Diane
Mack, Raymond W.
Macy, Michael W.
Madson, Richard
Maines, David R.
Manning, Wendy D.
Markham, William
Marshall, Harvey H.
Marshall, Susan E.
Martin, Jack
Marwell, Gerald
Marx, Gary T.
Massey, Douglas S.
McCarthy, John D.
McDaniel, Antonio
McFall, Stephanie L.
McLemore, S. Dale
McLeod, Jane D.
McPhail, Clark
Menard, Scott
Messner, Steven F.
Meyer, John W.
Michaels, James W., Jr.
Miethe, Terance D.
Minkoff, Debra C.
Mirowsky, John
Mizruchi, Mark S.
Model, Suzanne
Molm, Linda D.
Morgan, S. Philip
Mullan, Brendan
Mutchler, Jan E.
Myers, Martha A.
Nelsen, Hart M.
Nolan, Patrick D.
O'Brien, David J.
Ortega, Suzanne T.
Padavic, Irene
Page, Ann L.
Pagnini, Deanna L.
Pallas, Aaron M.
Palmer, Donald
Pampel, Fred C.
Parcel, Toby L.
Parish, William L.
Parnell, Allan
Paternoster, Raymond
Perrucci, Carilyn C.
Pinard, Maurice
Popielarz, Pamela
Postone, Moishe
Powell, Brian
Ranger-Moore, James
Reinharz, Shulamit
Roberts, Bryan R.
Rojek, Dean G.
Rosenberg, Janet
Roy, William G.
Rule, James B.
Rytina, Steven L.
Sakamoto, Arthur
Sanders, Jimy M.
Santiago, Anna
Schnaiberg, Allan
Schoen, Robert
Schulman, Michael D.

Schwartz, Barry
Seccombe, Karen
Sedaitis, Judith
Seeman, Melvin M.
Segal, David R.
Sheley, Joseph F.
Shehan, Constance L.
Shepelak, Norma J.
Sherkat, Darren E.
Shihadeh, Edward S.
Sigelman, Lee
Smith, David A.
Smith, Joel
Smith, M. Dwayne
Smith, Tom W.
Smith, Victoria
Snipp, C. Matthew
Sørensen, Aage B.
Sørensen, Annemette
South, Scott J.
Soysal, Yasemin
Spenner, Kenneth I.
Spitze, Glenna
Squires, Gregory
St. John, Craig
Stafford, Mark C.
Stahura, John M.
Steffensmeier, Darrell
Stevens, Gillian
Stinchcombe, Arthur L.
Sullins, D. Paul
Sullivan, Teresa A.
Sutton, John
Swicegood, C. Gray
Tausky, Curt
Thomson, Elizabeth
Tigges, Leann M.
Tiryakian, Edward A.
Tolbert, Charles M., II
Tolnay, Stewart E.
Tomaskovic-Devey, Donald
Trent, Katherine
Turk, Austin
Turner, J. Blake
Turner, Stephen Park
Useem, Bert
Villemez, Wayne J.
Walder, Andrew G.
Wallace, Walter L.
Warr, Mark
Wasserman, Ira M.
Waters, Mary C.
Watkins, Susan Cotts
Wellman, Barry
Western, Bruce
Wharton, Amy S.
White, Lynn K.
Whyte, Martin K.
Willer, David E.
Williams, J. Allen, Jr.
Williams, Rhys H.
Williams, Robin M., Jr.
Williamson, John B.
Wilson, Thomas C.
Winship, Christopher
Woodrum, Eric M.
Yamaguchi, Kazuo
Yinger, J. Milton
Zetka, James R., Jr.

SOCIAL FORCES

University of North Carolina Press
P.O. Box 2288
Chapel Hill, NC 27515-2288

Second class postage pai[d]
at Chapel Hill, N.C.
and additional post offices

RETURN POSTAGE GUARANTEED

SOCIAL FORCES

AN INTERNATIONAL JOURNAL OF SOCIAL RESEARCH
ASSOCIATED WITH THE SOUTHERN SOCIOLOGICAL SOCIETY

VOLUME 74: NUMBER 4 JUNE 1996

IN GENERAL

1. Mail manuscripts (5 copies) to: Social Forces, Department of Sociology, Hamilton Hall, University of North Carolina, Chapel Hill, NC 27599-3210.

2. The paper must not be submitted to another journal while under review by SF.

3. Enclose a $15 processing fee. Papers are not reviewed until this fee is paid. The fee is waived for papers solely by currently enrolled students.

4. Manuscripts are not returned to authors.

5. Authors transfer copyright to SF. Permission fees for reprint of articles are split between authors and journal.

PREPARATION OF MANUSCRIPTS

SF Form For review, papers may be submitted in the form used by SF, American Sociological Association or other major sociological journals. Authors of accepted manuscripts are responsible for putting them into SF form.

Format, Organization, and Style Submit 5 clearly legible copies on 8 ½ by 11 good quality white paper. Double space typing throughout. For an accepted paper being revised to fit SF style, leave at least 1" margins; do not justify margins if using a word processor.

Order Organize the manuscript in this order: cover page; abstract; text; endnotes; references; tables; figures.

Cover Page Give title; author(s); affiliation(s); and a footnote (*) indicating name and address of the author to whom requests for offprints or other correspondence should be sent ("Direct correspondence to ___") and acknowledgment (if any) of financial or other assistance.

Abstract On a separate page, preceding the text, write a summary, 125 or fewer words (70 or fewer for a Research Note).

Endnotes Use only for substantive comments, bearing on content. Number consecutively from 1, double space, and append on a separate page.

Tables Type each table on a separate page. Insert a location note — "Table 1 about here" — in the text. For substantive footnotes relating to a table or figure use lowercase letters. Use tabs to separate columns.

Figures Must be artist-drawn, camera ready. Do not send original copy with a manuscript submitted for review. For review, figures may be roughly drawn but they must be neat and legible. For footnotes to figures, use same symbols as for footnotes to tables.

Mathematical Symbols (and others likely to be obscure to compositor) Clarify with encircled words written in the margin.

References in Text Indicate sources as illustrated below:

- when author's name is in text — Lipset (1960); when author's name is not in text (Lipset 1960)
- use page numbers only for direct quotations or specific notes or table — (Braudel 1969:213)
- for more than 3 authors use "et al."
- with more than 1 reference to an author in the same year, distinguish them by the use of letters (a,b,c) with year of publication (1975a)
- earlier publication should precede later publication in brackets with parentheses (Tocqueville [1835] 1956)
- enclose a series of references — in alphabetical order — in parentheses, separated by semicolons (e.g., Adler 1975; Adler & Simon 1979; Anderson, Chiricos & Waldo 1977; Bernstein et al. 1977; Chesney-Ling 1973a, 1973b).

References Following Endnotes List authors alphabetically, by surname. Spell out first names of all authors and editors. For authors with more than one work cited, list works earliest to latest. For articles, next give title of article (caps and lower case), name of journal, volume number, and pagination. For books and monographs, give title, followed by publisher.

Format of References Please spell out the first names of all authors and editors, unless they use only their initials or a first initial and a middle name in the source cited (e.g., Paul Radin, T.S. Eliot, and J. Owen Dorsey).

Elder, Glen H. 1975. "Age Differentiation and the Life Course." Pp. 165-90 in *Annual Review of Sociology*. Vol. 1, edited by Alex Inkeles, James Coleman, and Neil Smelser. Annual Reviews.

Myrdal, Gunnar. [1944] 1962. *An American Dilemma*. Harper & Row.

Ritzer, George. 1975a. *Sociology: A Multiple Paradigm Science*. Allyn & Bacon.

___. 1975b. "Sociology: A Multiple Paradigm Science." *American Sociologist* 10:156-67.

Computer Files Authors analyzing data available in this format should cite their source, indicating producers and distributors as in these examples:

Louis Harris and Associates. 1975. Harris 1975 Nuclear Power Survey #2515 [computer file]. New York: Louis Harris and Associates [producer]. Chapel Hill: Louis Harris Data Center, University of North Carolina [distributor].

U.S. Bureau of the Census. 1979. 1970 Census of Population and Housing, Fourth Count Population Summary Tape [computer file]. Washington: U.S. Bureau of the Census [producer]. Rosslyn, Va.: DUALabs [distributor].

SOCIAL FORCES

JUNE 1996, VOLUME 74, NUMBER 4

ISSN 0037-7732 Library of Congress Catalog Number 24-31023

Second class postage paid at Chapel Hill, N.C. 27514 and additional mailing offices. *Postmaster: Send 3579 to Social Forces, UNC Press, P.O. Box 2288, Chapel Hill, NC 27515*

Claims for undelivered copies must be made within the month following the regular month of publication. The publisher will supply missing copies when losses have been sustained in transit and when the reserve stock will permit.

Social Forces is issued quarterly:
September, December, March, June.

Social *Forces* is published by the University of North Carolina Press for the Department of Sociology, with the assistance of the Institute for Research in Social Science, University of North Carolina at Chapel Hill.

Rev. 2/2/96

INFORMATION FOR *SF* SUBSCRIBERS

For Volume 75

Social Forces is issued quarterly: September, December, March, June. Subscriptions are accepted on a four-issue basis only. We recommend that readers subscribe for a volume year — beginning in September and ending with the June number. Subscriptions entered after the appearance of the first issue in a volume should specify whether the subscriber wants the current volume (including back numbers) or wants to begin the subscription with the next volume.

Subscription Rates

Individuals - U.S.

Regular:	
1 year:	$32
3 Years:	$91
ASA:	
1 year:	$25
3 years:	$70
Students:	
1 year:	$12

Institutions - U.S.

1 year:	$52
3 years:	$157

Single Copies - U.S.
US $14

Individuals - Outside U.S.

Regular:	
1 year:	US $40
3 Years:	US $115
BSA, CSAA, ASA (non-U.S.):	
1 year:	US $35
3 years:	US $98
Students:	
1 year:	US $20

Institutions - Outside U.S.

1 year:	US $64
3 years:	US $181

***Social Forces* Cumulative Indexes**
Volumes 1-50 (1922-1972): US $12
Volumes 51-58 (1972-1980): US $12

(For single copies and the Index add US $2.50 for mailing outside the U.S.)

Back Numbers. Reprinted Volumes 1 through 40 are available from the Johnson Reprint Corporation, 111 Fifth Avenue, New York, New York 10003. Back numbers of Volume 41 and subsequent issues are available from the managing editor of *Social Forces*, Institute for Research in Social Science, Manning Hall CB #3355, University of North Carolina, Chapel Hill, North Carolina 27599. The four most recent back volumes are available from The University of North Carolina Press. Volumes are available in microfilm from University Microfilms, 300 North Zeeb Road, Ann Arbor, Michigan 48106.

Abstracts and Indexes. The journal's articles are indexed or abstracted in *A.B.C. Political Science & Government, Abstracts in Anthropology, Academic Index, Adolescent Mental Health Abstracts, American Bibliography of Slavic & East European Studies, Applied Social Sciences Index & Abstracts, Automatic Subject Citation Alert, Chicano Periodical Index, Communication Abstracts, Criminal Justice Abstracts, Current Contents, Current Index to Journals in Education, Current Literature in Family Planning, Education Administration Abstracts, Excerpta Indonesica, Family Studies Abstracts, Geographical Abstracts, Historical Abstracts, Leisure, Recreation & Tourism Abstracts, Linguistics & Language Behaviour Abstracts, Middle East: Abstracts & Index, PAIS: Public Affairs Information Service, Population Index, Psychological Abstracts, Social Work Research & Abstracts, Social Sciences Index, Social Sciences Abstracts, Social Science Citation Index, Sociological Abstracts, Studies on Women Abstracts, Urban Affairs Abstracts, Urban Studies Abstracts, and World Agricultural Economics & Rural Sociology Abstracts, International Bibliography of Periodical Literature*, and book reviews in *Book Review Index.*

Change of Address. Notice of change of address should reach the publisher at least 6 weeks before publication of the next number. Please include both old and new addresses and zip codes. *Communications regarding subscriptions should be addressed to: Social Forces* Subscriptions, The University of North Carolina Press, Box 2288, Chapel Hill, North Carolina 27515.
Editorial and business communications should be addressed to: Social Forces, 168 Hamilton Hall, University of North Carolina, Chapel Hill, North Carolina 27599-3210.

World Wide Web Page. Internet users can access information regarding *Social Forces*, including advertising rate card, authors' guide, and subscription form, at: http://www.unc.edu/depts/irss.sfhome.htm.

SOCIAL FORCES

JUNE 1996, VOLUME 74, NUMBER 4

Contents

Social Forces 74:4, June 1996

BOOK REVIEWS

Who Fought in Vietnam?*

CYNTHIA GIMBEL, *Caliber Associates*
ALAN BOOTH, *The Pennsylvania State University*

Abstract

Using a sample of 4,338 Vietnam-era veterans, several factors indicating biosocial characteristics, race, mental aptitude, antisocial behavior in school, and military history were examined for their effects on (1) risk of combat arms assignment, (2) risk of assignment to Vietnam, and (3) level of combat exposure once in Vietnam. Also considered was the possible effect of general war strategy, as indicated by whether it was early or late in the war. Factors that indexed ability to perform noncombat support jobs had the most significant effect on combat arms assignments and assignment to Vietnam factors; these factors were also significant in predicting combat exposure. Biosocial predispositions to aggression and stress management are also associated with degree of combat exposure. Battlefield strategy also influenced the selection of individuals for combat.

During the Vietnam War, a soldier was not assigned to combât immediately on entry into the military. Instead, combat exposure was the end product of a three stage process that proceeded from job assignment, to assignment to Vietnam, and finally to combat. Assignments at each of these stages either increased or decreased a soldier's risk of combat exposure. These assignments were not entirely random, but rather depended on the characteristics of the individual soldier and the needs of the military at that time. Prior research on this topic has been driven primarily by hypotheses concerning discrimination and thus has focused on race and class as predictors of combat-related assignments or experiences. (See Appy 1993, Badillo & Curry 1976, Martin 1986, Moskos 1970, and Zeitlin, Lutterman & Russell 1973.) Exceptions are Card (1983) and Frey-Wouters and Laufer (1986). In addition, this earlier research examined stages of this process separately and did not account for the influence of predictors on previous stages. (An exception is Martin 1986.) Also, prior research examined issues of selection into combat for either the entire period of the war or for very limited time periods; dramatic changes in both strategic goals and public sentiment that occurred during the course of the Vietnam War have been ignored

** Direct correspondence to Cynthia Gimbel, Caliber Associates, 10530 Rosehaven St., Suite 400, Fairfax, VA 22030.*

 Social Forces, June 1996, 74(4):1137-1157

There are three important questions we would like to consider here. First, were there individual characteristics beyond race and class that increased a soldier's risk of combat by increasing the probability of receiving a combat-related job assignment, assignment to Vietnam, and ultimately, combat experience? Second, at what stage in the process did each of the individual characteristics matter? Third, did the influence of these characteristics vary by time period in the war?

Using a sample of 4,338 surviving Army nonofficers who served between 1965 and 1971, this research employs multivariate analysis to test two models of the influence of personal characteristics on assignment in the military. The potential combatant model states that men were placed in assignments that carried a greater risk of combat because they possessed qualities that would make them better combatants. The potential support staff model states that men were placed in assignments that carried a greater risk of combat because they had characteristics that indicated a lack of ability to handle complex support positions. Each model has associated with it a different labor queue and different variables that are operational at different points in time. We assess these models at each of the three phases in the process from induction to combat by statistically controlling for their influence on the previous phase. The difference in validity of these models according to the time period of the war is also assessed by examining the interactions between the soldiers' time of service and personal characteristics.

We focus on risk of combat exposure among survivors of the Vietnam War rather than all soldiers because our broader interests are in the long-term individual outcomes related to combat exposure. Repeatedly, Vietnam combat veterans have been found to have suffered more negative life outcomes than either other veterans or civilians of the same ages (Frey-Wouters & Laufer 1986; Kulka et al. 1991). These differences, for the most part, have been attributed to combat experience. However, this explanation ignores the life context in which combat occurs. If combat experience were randomly assigned to all men entering the military, then understanding this life context would be irrelevant. However, accepting that combat was not randomly assigned, then understanding its negative life outcomes cannot begin until we learn who was at greatest risk of combat exposure and whether the characteristics that identify these individuals contribute in any way to the later-life outcomes. By identifying characteristics which placed these men at differential risk of combat exposure, we can better understand the relationships between the experience of combat and the preceding and subsequent life experiences. In this study, we do not examine other important issues such as what characteristics predict induction into the military or what factors predict whether or not an individual was killed in combat. We suggest, however, that the variables which predict combat exposure would also predict who is killed in combat. Our study considers soldiers who served in the military and who survived the war and approximately fifteen years of post-military life.

THE MODELS

According to the theory of labor queues, organizations select individuals based on their ranked desirability for the task at hand (Thurow 1975). The military had two competing task queues: combat and support.

Those with greater combatant potential would be selected for the labor queue for combat. Our assumption is that indicators of biological and behavioral predispositions to aggression and good stress management are measures of desirability for combat tasks . Soldiers with these characteristic would not be the only ones who experienced combat but they might have been somewhat more likely to experience combat or to experience more combat than other soldiers.

For the support task queue (associated with noncombat jobs), those who had greater potential for handling complex support assignments would have been selected. While not all support positions involved complex tasks, selection for support could have been based on an assumed need for more complex skills than were required in combat. Desirable characteristics for support staff are mental aptitudes indicating trainability, commitment to the military (indicated by status as volunteer or draftee), and maturity (indexed by age at entry). (See Segal 1989 for a discussion of the military's reluctance to train low aptitude soldiers.)

The assignment of soldiers to a direct combat role was the outcome of a complex process composed of three phases: (1) receiving a combat arms assignment rather than a combat support assignment, (2) being sent to Vietnam rather than to Europe or remaining in the U.S., and (3) experiencing combat in Vietnam. The assignment made at each phase could be based on degree of desirability as a combatant, desirability as support staff, or both. The set of significant factors in the decisions could vary according to the phase since both the requirements and the available pool of personnel differed at each phase. At the outset, decisions were made Army-wide about the large groups of men who were available for assignment. Various characteristics such as test performance, age and draft status are expected to have been important factors because they were readily accessible. At the end of the process (in or near the field of combat), decisions were made at the company or platoon level about a few men. Here, personal qualities such as aggressiveness and ability to handle stress were more readily accessible and thus likely to be considered.

WHO RECEIVED COMBAT ARMS ASSIGNMENTS?

Combat arms assignments entailed direct confrontation with the enemy (e.g., light weapons infantry, armor crewman, field artillery crewman, automatic weapons crewman) and were made at the conclusion of basic training. Other noncombat assignments, such as combat support and combat service support, were made at the same time.[1] In 1968, only 12% of all servicemen were in combat jobs (Baskir & Strauss 1978). The other 88% comprised the complex organization of specialties backing combat soldiers, a significant number of which were in high-skill occupations. Evidence suggests that combat arms jobs were filled predominantly by men with lower mental aptitudes and skills (Baskir & Strauss 1978; Laurence & Ramsberger 1991) who were presumed unable to

handle complex support positions. This assignment policy held even as manpower demands increased. When more manpower was needed for combat roles, the noncombat soldiers transferred to combat often had low IQs (Baskir & Strauss 1978). Heavy demands for manpower also led to the lowering of aptitude and educational standards for entry into the military. During the Vietnam War, this occurred at least four times (Eitelberg 1988). Thus, we would expect that those scoring low on the Armed Forces Qualifying Test (AFQT) would have been more likely to get combat arms assignments than to receive assignment to support positions.

Draftees may not have been preferred for highly skilled support jobs because they were only in the service for 24 months and because they were presumed to be not committed to the military, since they did not volunteer. It was not cost-effective to invest heavily in training a draftee who would only use his skills for a short time. In addition, assigning draftees to combat may have aided in recruitment. One recruiting strategy was to offer relatively safe assignments to those who entered voluntarily for an extended period (Baskir & Strauss 1978). Preferential treatment also may have been given to volunteers in the form of noncombat assignments in the hopes of reenlistment.

A soldier's value to the military might also have been a function of age at entry. Young recruits were expected to have had poor academic performance. Soldiers aged 19 and younger may have appeared to the Army not to be worth the training investment. Perhaps they were viewed as too immature to follow complex instructions, be thorough, work autonomously, or take responsibility for tasks. Thus, they would not have been preferred for complex support positions.

These considerations lead us to the first hypothesis:

Hypothesis 1: Having a low intelligence/aptitude score, being young, and being drafted increased the probability of receiving a combat arms assignment.

If these variables are the primary predictors for receiving a combat arms assignment, then the potential support staff model explains assignment to combat arms better than the potential combatant model. Support for the hypothesis is also consistent with the idea that early large-scale screening of men was based upon readily available status characteristics and test results rather than on qualities that require direct observation to discern. On both counts, combat arms assignments would appear to have been received by default.

WHO WAS SENT TO VIETNAM?

While those with combat arms assignments were in great demand in Vietnam, Army combat personnel also were being sent to other overseas locations such as South Korea and Europe, as well as being kept stateside. Thus, scores on aptitude tests, type of entry, and age at entry may have figured in decisions about assignment to Vietnam. For example, men with high levels of aptitude or education, regardless of assignment, may have been kept stateside to work for officers who wanted reliable clerks or assistants (Baskir & Strauss 1978). Draft status may have affected later timing or location of assignment through its connection to reenlistment. Additional commitment to the military through early

reenlistment on the part of volunteers (spurred by impending assignment to Vietnam) may have insured new noncombat training and thus prolonged avoidance of the war zone for some (Baskir & Strauss 1978). Draftees, however, would have been less likely to reenlist than those who joined voluntarily the first time, and are thus less likely to have benefitted from this. Emotional and work-related maturity (indicated by age at entry) also may have kept some men out of Vietnam. Mature workers might have been preferred for tasks and training that would keep them in the U.S. because they were perceived as better prepared. These points lead us to the second hypothesis:

Hypothesis 2. Having a combat arms assignment, low intelligence/aptitude score, being young, and being drafted increased the probability of being sent to Vietnam.

With the exception of holding a combat arms assignment, evidence in favor of this hypothesis would support the validity of the potential support staff model for this phase of the selection process. Like receiving a combat arms assignment, being sent to Vietnam may have been a default assignment partly brought about by the requirements for massive screening.

WHO ACTUALLY FOUGHT?

The degree of exposure to direct combat in Vietnam was due to many factors which may have been outside either his or his immediate superior's control and thus independent of his personal attributes. For example, assignment to a particular area of Vietnam influenced the amount of combat exposure immensely; newly won areas often entailed more combat than areas of established strength. However, it is also very plausible that personal qualities had their strongest impact on decisions made in the actual field of battle.

Company commanders survived by selecting those soldiers whose personal characteristics yielded successful combat outcomes within Vietnam's specific environment. Leaders may have exposed certain individuals to high-risk combat assignments because they demonstrated behaviors that would increase the probability of victory or safety. Much like mid-level managers within a large corporation, the interests of company commanders focused on immediate accomplishment, in contrast to the higher-level executives (away from the battlefield) whose concerns are the long-term functioning of the larger organization (Scott 1981).

While a potential combatant should certainly be highly disciplined, confidence, determination, aggression, and boldness are also essential to the military mission (Bryant 1979). Individuals willing to take a certain amount of risk who were aggressive and did well under stress would have been desirable as combat soldiers. Therefore, according to the potential combatant model, individuals with these characteristics would have been exposed to more combat than soldiers without these characteristics.

We examine several indicators of aggressiveness and stress management: testosterone levels, cortisol levels, antisocial behavior in schools, and antisocial behavior in the military. Testosterone level is a useful indicator of aggressive tendencies. A number of studies have demonstrated a consistent and moderately strong relationship between testosterone and aggression in youth and adults

(e.g., Christiansen & Knussmann 1987; Ehrenkranz, Bliss & Sheard 1974; Olweus et al. 1980), antisocial behavior such as fighting, nontraffic arrests, and drug use in adults (Dabbs & Morris 1990), juvenile delinquency and adult deviance (Booth & Osgood 1993), and sensation-seeking (Daitzman & Zuckerman 1980. Testosterone has been shown to be stable over time, with reliabilities ranging from .50 to .65 for measures taken in the span of one day and up to six years apart (Dabbs 1990; Gutai et al. 1990). Therefore, if our models are correct, basal testosterone levels obtained long after combat should predict an increase in combat exposure.[2]

We draw on another biosocial variable, cortisol, as an indicator of ability to handle stress. While often studied as a response to stress (Henry 1992; Sapolsky 1992), basal levels of cortisol indicate an aptitude for dealing with stress. For example, research has shown that the most consistently successful competitors have lower cortisol levels. These individuals remain cool in times of intense competition (Booth et al. 1989) and are predisposed to handle stress well. This quality appears early in life. Kagan, Reznick, and Snidman (1988) have shown that the propensity to handle unfamiliar situations and engage other individuals appears in youngsters age seven and under and is related to low cortisol production. Research on soldiers in high-stress assignments in Vietnam (Green Berets and helicopter ambulance pilots) also shows uniformly low cortisol levels (Bourne, Rose & Mason 1967, 1968). Thus, men predisposed to handle high-stress situations well would have been more desirable in the high-stress atmosphere of battle and would have received appropriate assignments.

Once behavioral indicators of aggressiveness and the ability to handle stress are taken into account, the biological indicators should recede into the background. Unfortunately for purposes of this analysis, we only have two behavioral indicators of aggression. One is a premilitary indicator of aggressiveness and risk-taking in school. The second is an indicator of disciplinary problems while in the military.

We view the risk-taking, aggressiveness, and stress-handling behavior as factors that affect decision making about combat duty by those making daily field assignments. Our third hypothesis follows from these considerations:

Hypothesis 3. Higher testosterone levels, lower cortisol levels, more school problems, and problem behavior in the military increased levels of combat exposure when controlling for combat arms assignment.

Support for the third hypothesis would indicate that the potential combatant model holds for the last phase of the decision process. However, given the risks associated with combat, the potential support staff model may also be valid at this stage. A tactic used in Vietnam was to have ground troops move into positions that would draw the enemy out, i.e., using them as bait (Helmer 1974); once exposed, artillery and tactical aircraft were brought in to destroy the enemy. It was believed this could be a viable tactic because of the capability to deploy sophisticated medical support rapidly to treat battle injuries. Yet, the inherent danger of this tactic was still great enough that the Army probably continued to assign those who could not fill more technically demanding support slots to combat activity. Hence, our fourth hypothesis:

Hypothesis 4: Having a low aptitude, being drafted, and being less than 19 years of age at entry into the military increased levels of combat exposure when controlling for combat arms assignment.

RACE

The role of race is examined in each phase of the process from induction to combat. Race could have been a discriminatory factor in any of the assignments. However, it is important to note that most research neither supports the notion of disproportionate African-American service during the Vietnam War (Berryman 1988; Card 1983; Fligstein 1980; Kulka et al. 1990), nor the idea that African-Americans had higher levels of combat exposure (Card 1983; Frey-Wouters & Laufer 1986) or higher casualty rates (Badillo & Curry 1976). In fact, of the major studies encompassing information from all or most of the war years, only Card (1983) finds any sign of racial differences in service: a higher proportion of nonwhites than whites in the ground forces. This difference did not result in differing levels of combat exposure for whites and nonwhites.

These findings do not support those reported early in the war by the popular press (Binkin et al. 1982) and by Moskos (1970) using U.S. Department of Defense data for war years through 1968. Casualty rates suggest that beliefs about disproportionate minority involvement may have been true early in the war but that the proportion of African-American casualty rates declined over time (Appy 1993). Although later research does not support these early conclusions, the belief lingers that a greater than expected number of nonwhite men fought in Vietnam; therefore, it must be retested continually using different data and different models. Thus, the conflict between research and public opinion leads to our fifth hypothesis:

Hypothesis 5. Race was not a significant factor in receiving a combat arms occupational assignment, being sent to Vietnam, or receiving combat exposure.

PERIOD OF THE WAR

Thus far, our arguments imply that the Army was a closed system in which all decisions were driven by internal imperatives (Scott 1981). While the U.S. was present in Vietnam, dramatic changes occurred in the level of American involvement, wartime goals, and public opinion (Krepinevich 1986). It seems unrealistic to suggest that the changing external environment would have no impact on the selection of men for combat. However, no broad changes in Army strategy during the Vietnam War were made in response to the war's negative progress. By 1968, when it was apparent that conventional strategies were not successful, American policy already had switched to withdrawal rather than victory (Krepinevich 1986). Thus, as the war progressed differences would not be expected in soldiers' assignment to the combat arms or to Vietnam since such changes would reflect overt policy changes at an organizational level. However, the impact of the changing external environment both in the U.S. and on the battlefield was probably more noticeable at the company level where lessons learned in combat and subtle shifts in wartime goals could be translated more easily into action.[3] Early in the war when victory rather than withdrawal was the

goal, an effort might have been made to utilize soldiers whose personal characteristics made them better combatants regardless of their other skills or potential. In fact, prior to the Vietnam War an effort was made to channel low aptitude men away from combat assignments. (See Moskos [1973] for a discussion of this policy as applied to the integration of the Army.) Once ultimate withdrawal became the goal, the potential support staff model, which served to keep talented men away from battle, may have been implemented over the potential combatant model. These considerations concerning period of the war lead to our sixth hypothesis:

Hypothesis 6. Factors that enhance combat desirability (testosterone, cortisol, school problems, behavior problems in the military) will be stronger predictors of combat exposure prior to 1968 than in 1968 and after. Factors that decrease desirability as support staff (low AFQT scores, drafted, young at entry) will be stronger predictors of combat exposure in 1968 or later than they are prior to 1968.

Research Procedures

THE SAMPLE

The sample is composed of men who served in the Army between 1965 and 1971. There are no officers and the men performed only one term of service. The men were randomly selected from military records of all nonofficers who served in the Army from 1965 to 1971. They were a part of a Congressionally mandated study of the long-term psychological and physical consequences of service in Vietnam conducted by the Centers for Disease Control (CDC 1989a). The men were medically and psychologically examined between June 1985 and September 1986. Sixty percent of those selected took part in the study. The most common reason for refusal to participate in the extensive examinations was work-related (CDC 1989a). When those who participated are compared to those selected for study, we see only small differences in the two groups. Eighty-seven percent of the selected Vietnam veterans were white, while 82% of those who actually took part in the study were white. Twenty-six percent of those selected fell into the lowest category of the enlistment technical test, while 23% of the participant were in this category. No more than two percentage points separated those who participated from those who were selected with respect to the distribution on age at entry, whether they were drafted, tactical military specialty, other categories of the enlistment technical test, year of entry, pay grade at discharge, and honorable discharge. The modal difference is 1% (CDC 1989). While this sample does not permit conclusions about other branches of the military, officers, and those who did not survive or reenlist, it is representative of those who served in the Army, which is the largest service.

The first analyses were conducted on 4,338 of the examination participants who had no missing data. Because the sample overrepresented those who went to Vietnam, the data were weighted to achieve the proportion (21%) of personnel in Vietnam during July 1968 (Department of the Army 1968). Our second set of analyses were conducted on 2,389 of the participants who actually served in Vietnam.

These data do not allow us to examine the full range of soldiers who entered the Army and went to Vietnam. The respondents are selected for having survived the war and the ensuing years of peace. This concern is especially germane to the case of African-American soldiers. As discussed previously, early in the war race was a factor in combat deaths. However, for the entire war, "black casualties were only slightly disproportionate (12.5 percent from a population of 11 percent)" (Appy 1993:19). Thus, we may underestimate the effects of race in the course of our analyses but given the changes in proportion of African-American casualties over time, the bias, if any, will be quite small.

MEASUREMENT

Dependent Variables

Combat Arms Assignment is an occupational specialty to which the individual was assigned and not the organizational unit in which they were placed. Jobs included in the combat arms assignment are infantry, combat engineers, field artillery, and air defense artillery. It is represented as an indicator variable coded 1 if men were assigned to a combat arms position and 0 otherwise. Vietnam Assignment is represented as an indicator variable coded 1 if the men served in Vietnam and 0 otherwise.

Combat exposure is represented using a variable computed as the total score summed over 12 items with scores ranging from 0 to 4: 0 = never had the experience, 1 = rarely, 2 = sometimes, 3 = often, and 4 = very often. The range for this variable is 0 to 48. The items are:

(1) Part of a land or naval artillery unit that fired on the enemy;
(2) Flew aircraft over South or North Vietnam;
(3) Stationed at a forward observation post;
(4) Received incoming fire;
(5) Encountered mines and booby traps;
(6) Received sniper or sapper fire;[4]
(7) Unit patrol was ambushed;
(8) Engaged Vietcong in firefight;
(9) Engaged North Vietnamese Army in firefight;
(10) Saw Americans killed or injured;
(11) Saw Vietnamese killed or injured; and
(12) Killed someone or thought had killed someone.

Principal factor analysis indicates that this is a unidimensional scale with an alpha coefficient of .92.[5]

Biosocial Factors

Biosocial factors used are the hormone levels for testosterone and cortisol. Testosterone and cortisol concentrations were assayed from single serum samples collected at 8:00 a.m., before breakfast. A standard double-antibody radioimmunoassay was employed (CDC 1989b). For the weighted subsample of Vietnam-era veterans, testosterone levels ranged from 53 to 1,950 nanograms/deciliter. For the same group, the range of cortisol levels was 3.7 to 48 micrograms/deciliter.

Background Factors

Race is a dichotomous variable coded 1 for African Americans and 0 for all others.[6] We measure intelligence/aptitude using the men's scores on the Armed Forces Qualifying Test (AFQT), which is administered at the time of entry into the military. AFQT scores are categorized into four levels with 1 indicating the lowest scores on the test and 4 indicating the highest.[7]

School Problems

We use a four-item index composed of misbehaving in school, expelled or suspended from school, playing hooky, and fighting at school. Principal components analysis indicates that it is a unidimensional scale with an alpha coefficient of .60 and ranges from 4 to 8.

Military History

Three indicators of military history were used: age at entry, method of entry (draft or volunteer), and whether the men had behavior problems while in the military. A 1 for the respective variables represents age 19 or younger when entering the military, entry by the draft, and being absent without leave (AWOL) or confined while in the military.[8]

Method of Analysis

Weighted logistic regression is used to estimate the probability of receiving a combat arms assignment and assignment to Vietnam. Ordinary least squares regression is used to estimate the models of combat exposure. For consistency, in each model we test for the main effects of biosocial factors, background factors, school problems, and military history variables. For combat exposure, we estimate a series of five models entering blocks of variables one at a time: biosocial, background factors, school problems, military history, and combat arms assignment.

Findings

Zero order correlations, means, and standard deviations for variables used in these analyses can be found in Table 1. The majority of bivariate relations are in directions consistent with our hypotheses. The relationships between combat arms assignment and AFQT score, age at entry, and being drafted are all in the direction expected as are those between Vietnam assignment and AFQT score and age at entry. Similarly, the correlations for combat exposure are in expected directions.

With respect to race, no statistically significant relationship exists between race and combat arms assignment or being sent to Vietnam. However, a statistically significant positive bivariate relationship exists between race and combat exposure.

TABLE 1: Zero Order Correlations, Means, and Standard Deviations for Variables Used in Analysis[a]

	1	2	3	4	5	6	7	8	9	10	11
1. Testosterone		.05**	.04**	-.06**	.13**	.14**	-.07**	.12**	.05**	-.01	—
2. Cortisol	.03		.03	-.03	-.06**	.00	.04**	.00	.03	.01	—
3. Race	.03	.00		-.34**	.04**	.01	.03	.14**	.02	-.01	—
4. AFQT score	-.06**	-.02	-.31**		-.17**	-.19**	-.06**	-.19**	-.11**	-.09**	—
5. School problems	.08**	-.05**	.00	-.12**		.17**	-.09**	.21**	.05**	.00	—
6. Age at entry < 19	.08**	.02	-.05**	-.12**	.14**		-.25**	.14**	.06**	.05**	—
7. Drafted	-.02	.05**	.03	-.07**	-.08**	-.26**		-.13**	.09**	-.03	—
8. Problem GI	.04*	.02	.08**	-.11**	.18**	.10**	-.12**		-.01	-.02	—
9. Combat arms	.03	-.01	.01	-.06**	.00	-.02	.25**	.02		.10**	—
10. Vietnam	—	—	—	—	—	—	—	—	—		—
11. Combat	.04*	-.08**	.05**	-.15**	.17**	.08**	.06**	.09**	.49**	—	—
Mean (all, weighted)	680.39	18.14	.12	2.27	5.18	.47	.64	.10	.26	.21	—
Standard dev. (all, weighted)	174.65	4.17	.24	.67	.90	.38	.36	.23	.33	.31	—
Mean (Vietnam)	676.32	18.28	.11	2.14	5.19	.52	.62	.09	.35	—	16.42
Standard dev. (Vietnam)	236.98	5.54	.31	.86	1.19	.50	.49	.29	.48	—	10.73

[a] Upper half correlations are for the entire sample ($N = 4{,}338$) and are weighted to yield .21 in Vietnam. Lower half correlations are for those who went to Vietnam ($N = 2{,}389$) and are unweighted.

* $p < .05$ ** $p < .01$

COMBAT ARMS ASSIGNMENTS

When combat arms assignment is regressed on biosocial factors, background factors, school problems, and military history characteristics, the expected independent relationships emerge (Column 1, Table 2). The negative statistically significant relationship between AFQT score and combat arms assignment indicates that the probability of receiving a combat arms assignment increased by 21% with every quartile decrease in AFQT score, controlling for the effect of all other variables in the model. Entering the military under the age of 20 and being drafted also increased the probability of receiving combat arms assignments.

One unexpected finding emerged: testosterone has a positive effect on the probability of assignment to a combat arms position. This is surprising because it is not clear how Army decision makers could have information on testosterone or on related aggressiveness and risk-taking. The finding is also suspect because the testosterone-combat arms assignment coefficient is not influenced by the relevant behavioral measures of aggression. It is also a very small effect when

TABLE 2: Logistic Regression Coefficients of Combat Assignment and Vietnam Regressed on Biological Factors, Pre-Military Problems, and Military Factors While Controlling for Race and Intelligence[a]

	Combat Arms		Vietnam	
	Coef.	S.E.	Coef.	S.E.
Biosocial factors				
Testosterone	.0004	.0002*	-.0002	.0002
Cortisol	.01	.01	.00	.01
Background factors				
Race	-.10	.15	-.26	.17
AFQT Score	-.24	.06**	-.22	.06**
Pre-military problems				
School Problems	.06	.04	-.02	.04
Military history				
Age at entry < 19	.30	.10**	.18	.11
Drafted	.53	.11**	-.20	.11
Problem GI	-.22	.16	-.29	.18
Combat arms	—	.50	.11**	
Constant	-1.7298		-.6855	
-2 Log likelihood (full model)	2,772.725		2,511.948	
-2 Log likelihood (intercept)	2,842.634		2,559.251	
Pseudo R^2	.02		.01	
(N=4,338)				

[a] Data are weighted to yield .21 in Vietnam.

* $p < .05$ ** $p < .01$

or on related aggressiveness and risk-taking. The finding is also suspect because the testosterone-combat arms assignment coefficient is not influenced by the relevant behavioral measures of aggression. It is also a very small effect when compared to the size of the other significant coefficients in the model. There is only a 7% increase in probability of combat arms assignment for every standard deviation increase in testosterone level.

Despite this one unexpected finding, our first hypothesis is supported. Combat arms assignments were made primarily on the basis of factors indexing ability to perform noncombat support jobs (AFQT scores and age at entry) and long-term commitment to the military (draft).

ASSIGNMENT TO VIETNAM

When Vietnam assignment is regressed on biosocial factors, background factors, school problems, military history characteristics, and combat arms assignment, two of the expected main effect relationships emerge (Table 2, column 2). Low AFQT scores and obtaining a combat arms assignment each increase the probability of assignment to Vietnam. However, age at entry and draft have no effect, suggesting that more subjective inferences concerning maturity and commitment may not have played as great a role in assignment to Vietnam as other less subjective indicators.

The fact that the coefficient for being drafted is negative (although nonsignificant) is especially puzzling. Because the effect of draft on being sent to Vietnam may differ depending on combat arms assignment, we tested an interaction term between draft and combat arms assignment in our model and found it to be positive and statistically significant (analyses not shown). This indicates that being drafted and assigned to a combat arms job increased a soldier's probability of being assigned to Vietnam. In fact, further examination of the differences in probability of service for the four groups implied by the interaction term (drafted with combat arms assignment, volunteered with combat arms assignment, drafted with other assignment, and volunteered with other assignment) indicates that draftees with combat arms assignments were more likely to be sent to Vietnam than were draftees without such assignments, volunteers with combat arms assignments, and volunteers with other assignments. Being drafted did not mean automatic assignment to Vietnam but being drafted and assigned to combat arms placed men at great risk of this destination. To summarize, we find support for the potential support staff model because men with lower AFQT scores, combat arms assignments, and draftees with combat arms assignments were more likely to be sent to Vietnam. However, neither age at entry nor simple draft status seemed to play the expected part in this analysis.

COMBAT EXPOSURE

Combat exposure regressed on the biosocial factors reflects the relationships predicted by the potential combatant model (Table 3, column 1). Higher levels of testosterone are associated with greater combat exposure and lower levels of cortisol are associated with higher levels of combat exposure. The relationship between cortisol and combat exposure holds when the background factors of race and AFQT score are added into the model. The relationship between testosterone and combat exposure is only moderated slightly when the background factors are entered although it is no longer statistically significant (Table 3, column 2).

As expected in a multivariate model, race has no significant effect. The bivariate relationship between race and combat is accounted for by AFQT scores. The coefficient for AFQT score is significant and negative as expected, indicating that lower AFQT scores are associated with higher combat exposure. The fact that AFQT scores account for the bivariate relationship between race and combat exposure suggests that race alone was not a factor in exposing soldiers to

combat. However, it is possible that race modified the effect of AFQT. That is, African Americans who scored low on this aptitude test may have experienced greater combat exposure than whites with low scores. We tested this possibility by creating a race-by-AFQT interaction term and entering it into the equation. It was not statistically significant ($p = .29$).

When the school problems indicator is added to the model, the coefficient for testosterone is reduced dramatically (by 50%), but the cortisol coefficient is only reduced slightly. As these biosocial variables reflect predispositions to behavior, we expect them to recede in importance once the behavioral indicators are added to the equation. Clearly, school problems is a behavioral indicator of different levels of aggressiveness. Since direct behavioral indicators of pre-military ability to handle stress were not available in this study, the fact that the cortisol coefficient is not appreciably affected by the addition of the pre-military problems variable is not surprising. The coefficient for AFQT remains negative and significant.

Adding the military history variables (young at entry, drafted, and problem behavior while in the military) does not appreciably affect the coefficients for biosocial factors, background factors, or antisocial behavior in school. As expected, each of the military history variables has a significant effect on level of combat, independently contributing to the exposure level.

Adding combat arms assignment to the model causes two important changes: it adds considerably to the power of the model, increasing the R^2 from .06 to .29; and the coefficient for drafted drops from .09 to -.04. This suggests that once the effect of assignment on level of combat exposure is removed, draftees sent to Vietnam actually experienced slightly less combat than volunteers sent to Vietnam. It is important to note that even with combat arms in the model the coefficient for AFQT remains negative and significant.

TIMING OF SERVICE

Finally, understanding the combat process requires determining the nature of the interaction between the period of the war and individual characteristics, and the effect of this interaction on risk of combat exposure. Those who went to Vietnam were divided into cohorts based on initial date of Vietnam tour. In separate models, we test interaction terms formed by multiplying early service (1965-1967) and each of the model indicators: testosterone, cortisol, antisocial behavior in school, disciplined in the military, AFQT score, young age at entry, and drafted (analyses not shown). A significant negative interaction term was found for young, early Vietnam assignees and a significant positive term for the interaction between early service and AFQT scores. Among those who went to Vietnam early in the war, being young is associated with a decrease in combat score. In addition, the inverse relationship between AFQT scores and combat scores is not as strong for those in Vietnam in the early years as it is for those who went later on. These significant interaction terms demonstrate some changes in the conduct of the war depending on the war period. On the battlefield, the potential support staff model does not fit behavior early in the war as well as it fits behavior that occurred later when winning the war was out of the question.

TABLE 3: Standardized Regression Coefficients of Combat Regressed on Biological Factors, Premilitary Problems, and Military Factors While Controlling for Race and Intelligence[a]

	1	2	3	4	5
Biosocial factors					
Testosterone	.04* (.002/.0009)	.04 (.0016/.0009)	.02 (.001/.0009)	.02 (.0009/.0009)	.00 (.0002/.0008)
Cortisol	-.08** (-.16/.04)	-.08** (-.16/.04)	-.08** (-.15/.04)	-.08** (-.16/.04)	-.07** (-.13/.03)
Background factors					
Race		.01 (.20/.72)	.01 (.42/.72)	.01 (.47/.72)	.02 (.58/.62)
AFQT Score		-.15** (-1.83/.27)	-.13** (-1.58/.26)	-.11** (-1.36/.27)	-.09** (-1.12/.23)
Premilitary problems					
School problems			.15** (1.33/.18)	.14** (1.24/.18)	.14** (1.27/.16)
Military history					
Age at entry less than 19				.06** (1.27/.45)	.04** (.91/.39)
Drafted				.09** (1.97/.46)	-.04* (-.92/.41)
Problem GI				.06** (2.33/.76)	.04* (1.42/.66)
Combat arms assignment					.50** (11.20/.40)
Constant	.00** (17.93/.97)	.00** (22.22/1.15)	.00** (14.84/1.53)	.00** (13.10/1.58)	.00** (10.54/1.38)
R^2	.01	.03	.05	.06	.29
(N = 2,389)					

[a] Unstandardized regression coefficients/standard errors in parentheses.

* $p < .05$ ** $p < .01$

Discussion and Conclusions

Using a sample of 4,338 surviving Army nonofficers who served one term during the Vietnam conflict, the main effects of biosocial characteristics, background factors, antisocial behavior in school, and military history on (1) risk of combat arms assignment, (2) risk of assignment to Vietnam, and (3) level of combat exposure once in Vietnam were examined. Combat arms assignments were made primarily on the basis of factors that indexed ability to perform noncombat support jobs (AFQT scores and age at entry) and long-term commitment to the military (volunteered). Assignment to Vietnam depended on job assignment. In addition, a low AFQT score increased the probability of service in Vietnam; for draftees, a combat arms assignment further enhanced the probability of going to Vietnam. For the first two stages of the selection process, the analyses suggest that an individual's risk of combat was determined by his ability to perform jobs other than combat, i.e., support staff tasks. Moreover, even after occupational and locational assignments were made, the potential support staff model continued to operate, thereby influencing the level of combat exposure. Men serving in Vietnam with low AFQT scores and who were young at entry were more likely to see combat even when controlling for assignment. However, the potential combatant model also contributes to the explanation of combat exposure. Those who were aggressive and had good stress management skills were more likely to experience combat. Level of combat exposure was also somewhat influenced by whether the war was considered winnable. Those who had strong potential for support staff positions experienced more combat in the early stages than in the late stages of the Vietnam War.

In addition, our results support previous findings that racial differences did not exist in assignment or in combat. The role of race was examined for all three dependent variables. We find no direct effects of race on combat arms assignments, Vietnam assignment, or level of combat exposure. In addition, we find no indication of differing treatment for African Americans and whites based on AFQT score. However, given the importance of AFQT to our models and the significant negative bivariate correlation between race and AFQT we must consider institutional or nonpurposive discrimination as an alternate explanation for these findings. If African Americans were more likely to have low AFQT scores than whites, they also were more likely to see combat. Unfortunately, our data do not allow us to adequately develop multivariate models that predict AFQT score as an early stage in the process leading to combat. Our data also do not allow us to look at two other crucial aspects of the relationship between race and the Vietnam War: entry into the military as a dependent variable and socioeconomic status as a predictor variable. Examination of issues related to entry into the military would add to our understanding of the relationship between race and service in Vietnam by dealing with both issues relating to service avoidance and disproportionate rejection of certain groups because of low aptitude scores. (See Baskir & Strauss [1979] for a discussion of service avoidance and Segal [1989] for a discussion of low aptitude rejections.) As indicated by other research (e.g., Appy 1993), examination of socioeconomic

status as a predictor of combat-related experiences would add to our understanding of the links among combat, race and AFQT scores.

One variable in our analyses acted in an interesting way. While the relationship between behavior problems in the military and combat exposure is as predicted by the potential combatant model, the relationship between testosterone and behavior problems in the military in the combat exposure models suggests that our indicator of behavior problems in the military is not an indicator of aggressive behavior. Entering this variable into the regression equation does not mediate the relationship between testosterone and combat exposure or the relationship between antisocial behavior in school and combat exposure. This leads us to suspect that this indicator would better fit within the potential support staff model. One indication that individuals may be appropriate for support staff is their lack of behavioral problems while in the military. Being AWOL or in confinement appears not to indicate an aggressive person but perhaps to indicate a person who has problems with the authority structure of the military, an indicator of low potential for military success. Increased combat exposure for these problem GIs also could have been a form of punishment for past performance.

We also believe that our models, in general, may have portrayed these men as passive characters whose every movement within in the military is decided by others. This is not the case. The men were also actors who in many ways shaped their own destinies. Some soldiers may have reenlisted to avoid assignment to Vietnam as we pointed out earlier (Baskir & Strauss 1978). Others may have used presentation and negotiation skills to avoid combat once in Vietnam. Baskir and Strauss (1978) provide an illustrative example:

> In all, a soldier had a better-than-even chance of getting an assignment away from the search-and-destroy missions and other field operations. Anyone who made a good impression had a good chance of staying out of combat. As Pat Brannigan, now a career Army officer, recalled: "When I landed, I talked about computers to anyone who would listen. I knew that computers were always kept air-conditioned and far from combat areas. As it turned out that's how I spent my year." (59)

(See also Curry [1985] for a discussion of manipulation of the military personnel system.)

Through their own actions, individuals also may have become involved in situations that carried a high risk of combat in order to attain a sense of control (Appy 1990) or a psychological high (Appy 1990), or because of the opportunity to perform and be rewarded for skill and proficiency by receiving admiration from others (Appy 1990). Moreover, those with higher testosterone may have been predisposed to seek combat as an extension of their hormone-driven impulses. A feedback loop may have existed in which the performance decisions of individual soldiers fed into the decisions of those making daily combat assignments. The officer's decision to make further combat assignments then reinforced the individual's predilection to engage in combat. Unfortunately, our current data do not permit us to assess whether these experiences stemmed from individuals either acting or being acted upon, or some combination of the two.

These analyses make several contributions to our knowledge of the Vietnam experience. First, in our treatment of the determination of combat exposure as

a multiphase process, we have studied the importance of soldier characteristics at each. Intelligence and aptitude appear to be important factors in determining combat exposure at each stage of the process. There is a significant negative relationship between AFQT score and combat exposure for men who served in Vietnam even when controlling for the effect of AFQT score on combat arms assignment; this strongly supports the idea that even in the theater of war, the field of combat was the site for men not suited to other tasks. Our findings suggest that the decision-making strategies across all three phases of the process were the same. Previous research indicated the importance of intelligence/aptitude in this process but did not reveal its importance at each stage of the decision-making process.

Second, previously unidentified groups of men have been distinguished as having been at heightened risk for combat-related occupational assignment, assignment to Vietnam, and beyond that, actual combat exposure. The first group contains those not suitable for the training provided by military service, as indexed by aptitude/intelligence scores, draft status, and age at entry. The importance of aptitude scores in determining ultimate risk of combat exposure is profound, but the other status characteristics also play a role. The fact that this group had a heightened risk of combat exposure may provide an example of discriminatory results that were not rooted in prejudicial attitudes or purposely discriminatory behavior.

Another previously unidentified group is men with biosocial predispositions to aggressive behavior or good stress management. By identifying the importance of these characteristics in determining the level of combat exposure, we gain insight into the role of biosocial predispositions in social situations and into the way in which more personal characteristics may have determined an individual's risk of combat exposure. Our results suggest that while the decision-making process at the battlefield level was similar with regard to status characteristics, it was very different in its focus on men who could get the job done rather than on placing men in combat who could not do other jobs.

Third, the risks to the least advantaged soldiers appeared to increase when the battle was defined as lost. Despite efforts in 1967 and 1969 to reform the draft and thus democratize service in Vietnam (Appy 1993), this research provides evidence for a differential impact of personal characteristics on combat exposure depending on the time period of the war. We provide empirical support for the idea that talented men were channeled away from battle in the later years of the war. The declining overall support for the war appears to have influenced not only the number and location of soldiers who entered combat, but also the type of soldiers who actually fought.

Finally, our identification of the importance of status characteristics and biosocial factors in determining the ultimate level of combat exposure may help to explain why some individuals had problems following combat and others did not. The least able combatants may have suffered more distress after the war than those with biosocial predispositions to aggressive behavior or good stress management. The ability of the least able to deal with the aftermath of combat may have been severely hampered by the very lack of social and intellectual resources that brought them to combat. On the other hand, the combat-able or perhaps even combat-seeking soldier may have negotiated the combat experience

with less psychological trauma as a result of temperament or skills that may have made him a better combatant or combat-seeker. This temperament or these predispositions may have made combat a less stressful experience for him.

Further research in this area should focus on sorting out the reasons for the suffering experienced long after a war by some men and not by others and for the varying magnitudes of these problems. We believe that a focus on pre-military characteristics and experiences may provide part of the answer. First, biosocial predispositions may serve as protective factors against the trauma of combat. Second, the pathways from pre-military characteristics and experiences to later-life problems may be both direct and indirect through combat. Future research should focus on clarifying these paths. Understanding these relationships could have policy implications both for selection of individuals for future combat situations and for the care for those who return.

Notes

1. A support assignment did not prevent individuals from experiencing combat. For example, if someone with a service support position was located in an area overrun by the enemy, that individual was likely to be engaged in combat.

2. A full exploration of the role of testosterone in increased combat exposure must consider the possibility that basal levels of testosterone are the result, not the cause, of behavior (Kemper 1990). Extraordinary daily competitive pressures may result in higher basal levels of testosterone. Thus, it might be argued that repeated combat experiences might lead to higher basal testosterone levels even years after combat. However, studies of varsity tennis, wrestling, and community-level chess competition do not support this (Booth et al. 1989; Mazur et al. 1991). Even over the course of a season, postcompetition testosterone levels always return to precompetitive levels. There was no sign of a gradual increase, even among the most successful players. It is also possible that some men who served in combat found the risk very exciting (Appy 1990) and continued after the war to seek out high-risk activities to recreate this psychological high. These continued high-risk behaviors could possibly result in later-life elevated testosterone.

3. See Krepinevich (1986) with respect to counterinsurgency tactics.

4. A sapper is a soldier who tries to infiltrate a base or camp to destroy lives or property (Baker 1981).

5. This self-report combat experience measure comes from the Legacies of Vietnam Study (Centers for Disease Control 1988).

6. Analyses including indicator variables for white, African-American, Hispanic, and Native American yield nonsignificant coefficients for all subgroups. For reasons of parsimony, only the African-American indicator variable remains in the equation.

7. While AFQT scores are typically reported in categorical form, this form of coding for AFQT scores is not usually used. Typically, these scores are reported as five categories with the highest aptitude soldiers in category 1 (i.e., Laurence & Ramsberger 1991). For our analyses, the two lowest categories were collapsed into one because of low frequencies in category 5. Then we reversed the coding so that it would be appear logical to those unfamiliar with traditional methods of AFQT reporting.

8. Confinement was one form of nonjudicial punishment that could be ordered by local commanders in response to minor offenses against the Code of Military Justice or in response to violations of other military rules (Bryant 1979:30).

References

Appy, Christian G. 1993. *Working Class War: American Combat Soldiers and Vietnam.* University of North Carolina Press.

Badillo, Gilbert, and G. David Curry. 1976. "The Social Incidence of Vietnam Casualties." *Armed Forces and Society* 2:397-406.

Baker, Mark. 1981. *Nam: The Vietnam War in the Words of the Men and Women Who Fought There. William.* Morrow & Company.

Baskir, Lawrence, and William Strauss. 1978. *Chance and Circumstance: The Draft, the War, and the Vietnam Generation.* Knopf.

Berryman, Sue E. 1988. *Who Serves? The Persistent Myth of the Underclass Army.* Westview Press.

Binkin, Martin, Mark J. Eitelberg, Alvin J. Schexnider, and Marvin M. Smith. 1982. *Blacks in the Military.* The Brookings Institution.

Booth, Alan, and D. Wayne Osgood. 1993. "The Influence of Testosterone on Deviance in Adulthood: Assessing and Explaining the Relationship." *Criminology* 31:93-117.

Booth, Alan, Greg Shelley, Allan Mazur, Gerry Tharp, and Roger Kittok. 1989. "Testosterone, and Winning and Losing in Human Competition." *Hormones and Behavior* 23:556-71.

Bourne, Peter G., Robert M. Rose, and John W. Mason. 1967. "Urinary 17-OHCS Levels: Data on Seven Helicopter Ambulance Medics in Combat." *Archives of General Psychiatry* 17:104-110.

____. 1968. "17-OHCS Levels in Combat: Special Forces "A" Team Under Threat of Attack." *Archives of General Psychiatry* 19:135-140.

Bryant, Clifton D. 1979. *Khaki-collar Crime: Deviant Behavior in the Military Context.* Free Press.

Card, Josefina. 1983. *Lives after Vietnam: The Personal Impact of Military Service.* D.C. Heath.

Centers for Disease Control Vietnam Experience Study. 1988. "Health Status of Vietnam Veterans: I. Psychosocial Characteristics." Journal of the American Medical Association 259:2701-2707.

____. 1989a. Health Status of Vietnam Veterans: IV. Psychological and Neuropsychological Evaluation. Centers for Disease Control.

____. 1989b. Health Status of Vietnam Veterans: Supplement A. Laboratory Methods and Quality Control. Centers for Disease Control.

Christiansen, K., and R. Knussmann. 1987. "Androgen Levels and Components of Aggressive Behavior in Men." *Hormones and Behavior* 21:170-80.

Curry, G. David. 1984. *Sunshine Patriots: Punishment and the Vietnam Offender.* University of Notre Dame Press.

Dabbs, James. 1990. "Salivary Testosterone Measurement: Reliability Across Hours, Days, Weeks." *Physiology and Behavior* 48:83-86.

Dabbs, James, and R. Morris. 1990. "Testosterone and Antisocial Behavior in a Sample of 4,462 Men." *Psychological Science* 1:209-11.

Daitzman, Reid, and Marvin Zuckerman. 1980. "Disinhibitory Sensation Seeking, Personality and Gonadal Hormones." *Personality and Individual Differences* 1:103-10.

Department of the Army. 1968. "Strength of the Army (U): Part 1 Strength of the Active Army, 31 July 1968." Office of the Deputy Chief of Staff for Personnel in Washington, D.C.

Ehrenkranz, Joel, Eugene Bliss, and Michael Sheard. 1974. "Plasma Testosterone: Correlation with Aggressive Behavior and Social Dominance in Man." *Psychosomatic Medicine* 36:469-75.

Eitelberg, Mark J. 1988. *Manpower for Military Occupations.* Human Resources Research Organization.

Fligstein, Neil D. 1980. "Who Served in the Military: 1940-1973." *Armed Forces and Society* 6:297-312.

Frey-Wouters, Ellen, and Robert S. Laufer. 1986. *Legacy of a War: Theory, Research, and Treatment.* Sharpe.

Gutai, James, Wanju Eai, Ronald LaPorte, and Lewis Kuller. 1986. "The Reliability of Plasma Sex Hormone Measurements in Men for Epidemiologic Research." Unpublished manuscript. Wayne State University, Detroit, Michigan.

Helmer, John. 1974. *Bringing the War Home: The American Soldier in Vietnam and After.* Free Press.

Henry, James. 1992. "Biological Basis of the Stress Response." *Integrative Physiological and Behavioral Science* 27:66-83.

Kagan, Jerome, J. Steven Reznick, and Nancy Snidman. 1988. "Biological Bases of Childhood Shyness." *Science* 240:167-71.

Kemper, Theodore. 1990. *Social Structure and Testosterone.* Rutgers University Press.

Krepinevich, Andrew F., Jr. 1986. *The Army and Vietnam.* Johns Hopkins University Press.

Kulka, R.A., W.E. Schlenger, J.A. Fairbank, R.L. Hough, K.B. Jordan, C.R. Marmar, D.S. Weiss, and D.A. Grady. 1990. *Trauma and the Vietnam War Generation.* Brunner/Mazel.

Laurence, Janice H., and Peter F. Ramsberger. *Low-Aptitude Men in the Military: Who Profits, Who Pays?* Praeger.

Martin, Randy. 1986. "Who Went to War." Pp. 13-24 in *The Vietnam Veteran Redefined: Fact and Fiction,* edited by Ghislaine Boulanger and Charles Kadushin. Erlbaum.

Mazur, Allan, Alan Booth, and James Dabbs. 1991. "Testosterone and Chess Competition." *Social Psychology Quarterly* 55:70-77.

Moskos, Charles C., Jr. 1970. *The American Enlisted Man: The Rank and File in Today's Military.* Sage.

____. 1973. "The American Dilemma in Uniform: Race in the Armed Forces." *The Annals of The American Academy of Political and Social Science* 405:94-106.

Olweus, Dan, A. Mattsson, Daisy Schalling, and Hans Low. 1980. "Testosterone, Aggression, Physical and Personality Dimensions in Normal Adolescent Males." *Psychosomatic Medicine* 42:253-69.

Sapolsky, Robert. 1992. "Neuroendocrinology of Stress-response." Pp. 287-324 in *Behavioral Endocrinology,* edited by J.S. Becker, S. Breedlove, and D. Crew. MIT Press.

Scott, W. Richard. 1981. *Organizations: Rational, Natural, and Open Systems.* Prentice-Hall.

Segal, David R. 1989. *Recruiting for Uncle Sam: Citizenship and Military Manpower Policy,* University of Kansas Press.

Thurow, Lester C. 1975. *Generating Inequality: Mechanisms of Distribution in the U.S. Economy.* Basic Books.

Zeitlin, M., K.G. Lutterman, and J.W. Russell. 1973. "Death in Vietnam: Class, Poverty and the Risks of War." *Politics and Society* 3:313-28.

Announcement

The Foundation for the Promotion of Social Science Research on World Society — World Society Foundation — funds selected proposals for research on the structure of and changes in world society. The next deadline for applications is June 30, 1997. Selected projects may start in January 1998.

The funding activity of the Foundation also results from the contributions contained in the three volumes of the series *World Society Studies,* published by Campus (Frankfurt 1990) resp. Transaction Publishers (1992 & 1994).

Further detailed information including the application form is available at the following Web Site: http://www.unizh.ch/wsf. Printed documentation may be ordered at the following address:

World Society Foundation
c/o Sociological Institute
University of Zurich
Rämistr. 69
CH-8001 Zurich
Switzerland
FAX 41 1 252 10 54

The Toggle Switch of Institutions: Religion and Art in the U.S. in the Nineteenth and Early Twentieth Centuries*

JUDITH R. BLAU, *University of North Carolina at Chapel Hill*

Abstract

An institution is taken for granted when it is enveloped within a cultural and moral framework that is not contested. The usual account of the institution of high culture that emerged in the U.S. in the late nineteenth century relates to the growing importance of elites and the uncertainties of modernization. This raises an important puzzle: how did the visual arts that had been so dominated by the morality of popular religion acquire secular worth? Weber's writings provide a point of departure for analyzing institutional transformations involving art, religion, and the emergence of the avant-garde. A framework for understanding institutional change is sketched out. It includes the importance of ideology, social and economic interests, and intellectuals who articulate emerging new views. The metaphor of the "toggle switch," adapted from Weber's term, switchmen, captures the idea that transformation occurs when there are underlying cultural contradictions and the resolution of these contradictions depend on the investments and prospects of competing interests.

A prevailing view of art is that it is part and parcel of culture, and, as such, has a functional relationship with a set of values that underlies social life and large-scale institutions. Art, according to this view, exhibits considerable inertia. It does so because its development and legitimation are profoundly implicated in group life (Becker 1982), social class (Bourdieu 1984; Gans 1974), organizations (Blau 1989; Peterson 1976), or elite interests (DiMaggio 1982).

This perspective is consistent with institutional theory, which emphasizes that culture, generally, has a high degree of coherence and consistency (Parsons

* *This work accompanies quantitative research on the transformation of cultural institutions supported by the National Science Foundation (SES-9108923). Early versions of this article were presented and discussed at the annual meetings of the Fourth National Sociology Conference in Norway (1993), at Pennsylvania State University (1994), and at Vanderbilt University (1994). For their generous and helpful suggestions I thank Arild Danielsen, Jennifer Glanville, Richard Peterson, Alan Sica, Joseph Soares, Jonathan Turner, and an astute anonymous reviewer.* *Direct correspondence to Judith R. Blau, Sociology Department, University of North Carolina, Chapel Hill, NC 27599-3210.*

1951), is interwoven with social arrangements that are keyed by rules (DiMaggio & Powell 1983), and that group and symbolic boundaries are defined in terms of one another (Bourdieu 1991; Douglas 1986; Lamont 1992). In other words, there are functional relationships and congruences between culture and the social order, whether culture operates from the top down (Douglas 1986) or the bottom up (Becker 1982), whether it fosters social order (Parsons 1951) or is oppressive (DiMaggio & Powell 1983).

The difficulties with the conception of art as coherent with the social order is that elements of a taken-for-granted institution can be antithetical to one another. Because legitimation entails great tenacity of investments and interests, institutional contradictions often remain obscure while embodying potential for significant transformation. The present article is a case study of the dislocations between religion and art in the U.S. in the nineteenth century, and I pose the problem of how secular art became autonomous after decades of being subordinate to popular religion.

To account for the mechanism of institutional transformation, I draw on Weber's (1946b) metaphor of the "switchmen" (280) ("Weichensteller der Bahnen"; Weber 1920:252) of social change. According to Weber, the switchmen that alter the direction of social change involve "world images" that are created by ideas and are driven by interests. Elsewhere, he suggests that the switching is triggered by deep-seated cultural contradictions. Weber ([1930] 1976) traced cultural (religious) beliefs to the "needs" or the interests of particular groups, but intellectuals, the bearers of rational culture, play particularly important roles. Moreover, he suggests that art can function much as traditional religion does; that is, through its subjectivism, art provides an escape from the very disenchantment that rationalization and science produces (Weber 1968,1:607-10). Weber's writings on the relations between art and religion, and the role of intellectuals in social change, are sketchy at best, but in this case study of the Arnoldian movement in America, it is possible to provide an empirical account of the transformations in the relations between religion and art in the U.S. and in the way that intellectuals provided a language that led to the subsequent development of the avant-garde.

The basic assumptions that can be gleaned from Weber's writings are the following: (1) that ideas and worldviews have considerable autonomy and are consequential (Weber 1946a, [1930] 1976); (2) that worldviews have their origins in socioeconomic conditions because groups have different interests (Weber 1946a; also see Bourdieu 1984; Sadri 1995); (3) that there are critical moments when the contradictions among worldviews can most powerfully motivate social change, as Weber's (1946b) "switchmen" metaphor suggests; (4) that art, which emergences as a potent force under conditions of rationalization, competes with salvation religion (Weber 1946a, 1968, 1); and (5) that intellectuals acquire unusual visibility and persuasiveness during periods of great flux (see especially Sadri 1995; also see Swidler 1973).

Further Considerations: The Toggling of Institutional Change

Weber's own metaphor, "switchmen," implies a mechanism that simply alters the direction of change, and it is inconsistent with his insights about the conditions of cultural contradiction. I use instead the "toggle switch" as an apt metaphor. The toggle translates one small input into a far larger effect, and this effect is indirect and often not obvious. In this sense, the toggle imagery captures the complex instrumentalities of historical change and problematizes the character of change.

Some of the inconsistencies and vagueness in Weber's arguments can be resolved by explicitly using the concept of ideology (Bourdieu 1991), specifically that differences in interests and status express themselves in discourse, representation, and social space. In this way we can begin to unpack the conditions under which institutions acquire a taken-for-grantedness, to suggest that they do so in terms of ideology, and also to suggest that taken-for-grantedness masks strategies of containment (Jameson 1981) and conceals from participants their variances and contingencies (Adorno 1973).

This case study suggests that art and religion "and, by extension, other domains of culture" have themselves transformative relations that are not completely reducible to groups, classes, and organizations, but rely on them, as interests are expressed in ideological terms (Eagleton 1991; Tarrow 1994; Touraine 1988), and are embedded within the expansive worldviews provided by science, theology, and philosophy.

I shall argue that the prevailing ideology of salvation religion during the nineteenth century was pervasive, as it was highly compatible with economic liberalism (Weber [1930] 1976) and dictated the character of nation building (as evident, for example, in school and reform movements) and the rules for acceptable social behavior. Yet the high culture movement that was so antagonistic to popular religion drew on social Darwinism to legitimize social and economic hierarchies, and to be legitimized by them. A practical instrument for this transformation in cultural institutions was the Arnoldian movement in the 1880s and 1890s, successful perhaps, in part, because of the fact that Matthew Arnold was a credentialed foreign (that is, exotic) intellectual, but also because he appealed to class and status interests as he defended both the high culture movement and social Darwinism. This is not to credit Arnold with this important transformation, but rather to consider the Arnoldian movement as providing clarity for the articulation of emergent conceptions about religion and art in the social order.

High Culture and Elite Sponsorship: The Conventional Account

The prevailing interpretation of the emergence of the fine arts in America at the end of the nineteenth century is that elites drew on the distinction between popular and high culture to bolster class differences (Archer & Blau 1993; Blau 1991; DiMaggio 1982; Horowitz 1976; McCarthy 1982; Miller 1966; Minihan 1977).

At the end of the century, the entrepreneurs of high culture responded to the destabilizing effects of immigration, vice, pauperism, and industrialization while providing forms of class security for the middle class and opportunities of ostentatious display for the urban rich. As Trachtenberg (1982) notes, "Changes in social structure, the polarization of rich and poor and the growth of a salaried middle class anxious about its own status opened the opportunity — indeed, created the necessity — for the healing properties identified with high culture" (145).

Exclusionary symbols helped to map the terrain (Whitehall 1970), while the rejection of (vulgar) popular culture provided reassurance of "good taste," elegance, and distinction (Bourdieu 1984; Lamont 1992; Slater 1976). Also accompanying this account of elite sponsorship is an analysis of the organizational means by which art would be sacralized. Such organizations as museums and orchestra halls contributed to the institutionalization of high art, for without fashionable public arenas, claims to superior taste were suspect as being snobbish, trivial, and undemocratic (DiMaggio 1992; Hall 1984; Wiebe 1967).

DIFFICULTIES IN THE CONVENTIONAL ACCOUNT OF HIGH CULTURE

This interpretation of the role of elites and organizations in the development of high culture is misleading for the U.S. because it misspecifies the relations between religion and popular culture and ignores how the fine arts were defined in America throughout the nineteenth century. To be sure, the performing arts and, to some extent, literature, were thoroughly integrated into secular leisure (Barth 1980; Cmiel 1990; Kasson 1990; Levine 1988; Toll 1976). Levine (1988), for example, demonstrates that what was later considered to be high culture, such as Shakespearean plays and Italian operas, were widely popular and enjoyed along with vaudeville and fairs. However, this is not the case for the visual arts, which were bracketed through most of the nineteenth century and were governed by religious norms (Commager 1950; May 1976; Miller 1965; Perry 1989). Popular culture "entertainment" could be contained (Jameson 1981) within an ideological order of austerity, asceticism, and piety because entertainment did not constitute a worldview with potent moral meaning (Shi 1985).

Religious traditions had a secure hold on art, and legitimization of the visual arts, independent of religion, required an effective displacement from the rhetoric of religious morality to the rhetoric of aesthetic value. If there was a drift away from religious morality and asceticism at mid-century, the Civil War brought a renewed celebration of piety, spartanism, and the denunciation of effeminate and indulgent practices (Shi 1985).

Rapid industrialization after the war spurred increasing differentiation between city life and rural life and growing economic disparities between workers and owners as well as between the affluent elite and the growing middle classes. It was later, around the 1880s, that the number of immigrants expanded dramatically, and, at least as important, their origins were no longer from predominantly, or even largely, Protestant nations. Increasingly, status rested less on behavior that could be observed in communities, and more on secular and universalizing conceptions of worth, such as education, life-style, and wealth (for this general point, see Lockwood 1992:87-93).

Thus, economic elites, under the conditions of growing inequalities, played a major role in creating the conditions for the development of high culture in America. Yet the connections between religion and art had to be severed for high culture to emerge. These antithetical relations were deeply rooted in a particular conception of morality, and high art required ideological justification.

Normal Morality and American Exceptionalism

The nineteenth-century American bourgeois shared with their northern European counterparts deep mistrust of sexuality, sin, and evil. The Victorian sensibility "defined itself largely through its fixation on the dangers of desire" and through its at least nominal adherence to a code of "earnestness and self-denial" (Herbert 1991:30). This sensibility, according to Gay (1986), reflected a great preoccupation about the line between the pure and the impure. In America, the distinctions were amplified at the pulpit and in the Sunday school and sectarian conceptions of morality played an important role in people's lives. Yet a major contrast is that in America these lines were drawn around religious and ethnic differences, whereas in Europe class differences played the more important role.

Lacking a class framework for moral evaluations, religious differences were the dominant ones, and these were fine-tuned as the difference between evangelical denominations — Methodists, Baptists, Disciples — and established ones — Episcopalians, Presbyterians, and Unitarians (Blumin 1989). Foreigners were amazed at the level of piety in America, and accounts abound with descriptions of religious fervor (Ratzel [1876] 1988; Tocqueville [1836] 1945; Trollope 1832; also see Morgan 1973). Over the course of the century, more powerful means of social distance and moral evaluation were made in terms of a combination of ethnicity and religion, so that sectarian and established churches contended against the Catholic church and other immigrant churches over the proprietorship of existing social and cultural arrangements (Blau, Redding & Land 1993).

It would appear that the provisions of the First Amendment would dilute the role of religion in secular matters. Wills (1990) argues that, paradoxically, the separation of church and state enabled an intense religiosity (also see Demerath & Williams 1992). Thomas Jefferson, a Deist who had penned the First Amendment, contended that a "wall of separation" between church and state would actually enhance opportunities for religious mobilization: "free argument, raillery and even ridicule will preserve the purity of religion" (1950:549). Similarly, Madison justified separation in that an established state church would "retard Christian evangelizing" (1910:306).

Varieties of Morality

Differences in doctrine between the established churches and the sectarians were reflected in different conceptions of morality and reform. Established churches had somewhat formal church practices and shared a theology that stressed the uncertainty of salvation. Established denominations more readily

viewed church matters as coordinate with state interests and increasingly withdrew from direct involvement in social reform and charity efforts, relying instead on centrist approaches to reform, such as civic projects and legislation. In contrast, the sectarians had relatively communal and decentralized church bodies and embraced beliefs in salvation. Sectarians conceived of grace as attainable and, placing great emphasis on the community of believers, viewed reform efforts to be a matter of helping individuals in seeking redemption (Hutchison 1989; Lipset & Raab 1978; Niebuhr 1929; Troeltsch 1931).

Moral reform throughout the nineteenth century largely reflected these divisions between the established churches and sectarians. Early sectarian social reform movements in the 1830s defined issues in individualistic terms, and amelioration, not regulation, was the approach to social problems such as liquor and prostitution (Boyer 1978). In contrast, the American Moral Reform Association, and later the National Reform Association, were largely supported by established churches and emphasized "regulationist" or legalistic strategies, working through local and state legislative bodies and police departments (Baltzell 1964; Handy 1991; Pivar 1973).

Increasingly, the regulationist reform agenda of the established churches came to prevail over the more individualistic reform strategies of evangelicals, leading to the adoption of antivice legislation in the mid-1870s (Beisel 1993). Antivice campaigns had been waged throughout the nineteenth century, but never before had these campaigns been so infused by interventionist logic and legalistic strategies (King 1989; Timberlake 1963; Tyack & Hansot 1982). It was in this particular historical context that Matthew Arnold "an ardent establishmentarian and advocate of interventionism" arrived in the U.S. with considerable fanfare. Before describing the significance of his emblematic visits, it is necessary to indicate that urban Americans had earlier accepted fine arts traditions but repudiated them in the early decades of the nineteenth century.

Nineteenth-Century Arts

In contrast with the U.S., the visual arts flourished in Europe throughout the nineteenth century (Holt 1979, 1981). After 1793 a salon was held almost annually in Paris that was open to artists of all nationalities; there were regular public exhibits in England, Berlin, and Rome during the first half of the century; and after 1850 there were major art exhibits in many other cities as well, including Florence, Munich, Naples, and Brussels. In virtually every European country, art was fully institutionalized, supported by the state and accompanied by a dense network of artists' associations, critics, publications, private galleries, and buyers. Art was not, as is sometimes claimed, only for private collectors. The Louvre was opened in 1793 as a public museum and throughout the nineteenth century exhibits and competitions were public events, and art galleries in the largest cities, according to Holt's (1979, 1981) accounts, did a brisk business in sales. Although it is the case that art exhibits and sales were predominantly in the private sector and most public museums were founded after around 1850 (Blau 1995), the visual arts were as much taken for granted among the urban bourgeoisie as music, literature, and drama.

Americans, quite in contrast, were indifferent to the visual arts, if not overtly hostile, as their attitudes were described at the time by artists (Dunlap 1834; Green 1966; Jarves [1864] 1960). A telling indicator of this is the large number of American artists who spent most, or all, of their time abroad. This includes virtually all the currently acclaimed nineteenth-century American artists: John Copley, John Trumbull, John Vanderlyn, Washington Allston, Samuel Morse, Thomas Cole, Henry Inman, George Bingham, Winslow Homer, William Page, George Inness, William Chase, William Morris Hunt, Thomas Eakins, Francis Millet, Mary Cassatt, James Whistler, and John Singer Sargent.

This antagonism to secular painting did not, in fact, dominate in colonial and early postcolonial America. Painting was a reputable career and possessing works of art was not uncommon in New England through the end of the eighteenth and in the early decades of the nineteenth century, and there was a significant effort made by a small minority of artists and patrons to fortify American artistic traditions in art associations in major northeastern cities (Harris 1966). However, public opinion was becoming increasingly suspicious of aesthetic and literary tastes unchecked by plain standards of religious self-denial (Cmiel 1990; Kasson 1990). Around 1800 John Adams failed in his efforts to persuade Congress to pass laws promoting the arts and literature (Cummings 1991) and an 1815 congressional declaration stated that fashionable dress, intellectuality, leisure, and art were "foppish instruments of the devil" (Bruchey 1965:198).

The origin of this surge of what Bruchey (1965) terms "puritanicalness," at the beginning of the nineteenth century, was, according to many American historians (May 1976, 1983, 1991; Miller 1965; Nye 1960; Perry 1989; Shi 1985), the manifestation of the Second Awakening from around 1795 to 1835. "Puritanicalness" thoroughly rent the social fabric. It was responsible for a precipitous decline in intellectual life in the first decades of the nineteenth century (May 1976), and accompanied a dramatic and sudden adoption of vernacular and informal American English (Cmiel 1990). The rejection of English and Continental models, notably establishment religion as well as intellectual and artistic traditions, was consistent with the evangelical tenet that religion was a "matter of the heart" (Shi 1985; Walters 1978).

Artworks after the 1820s were increasingly scrutinized within this new moral framework. Evaluative criteria shifted to spirituality and religious sentiment, and the "painterly" traditions that continued to dominate in English and Continental art were rejected. Italian and French art was suspect for such obvious reasons as its nudity and traditional patronage by the Catholic church and the aristocracy, but even English landscape paintings and portraits, which in postcolonial America had been highly regarded, were viewed with suspicion by the American public and collectors. Those U.S. artists who continued to paint in the secular tradition of their English contemporaries — such as Constable, Millais, Hunt, and Turner — were virtually ignored by American collectors (Harris 1966), and, as already noted, there was a sustained migration of artists to England, France, and Italy.

Artists themselves were frustrated by the displacement of aesthetic values by religious ones. Dunlap (1834) complained that in the many cities he toured with his paintings, local clergy would be called upon to evaluate and judge his

works for their religious worth. By the middle of the century, artists began to accept these conditions and along with writers justified their work under the broad umbrella of transcendentalism (Shi 1985). In an article in the *Crayon,* Durand (1855:97) admonishes, "it is only through the religious integrity of motive by which all real Artists have ever been actuated."

Art historian Joshua Taylor (1976) describes the widespread influence of the Society for the Advancement of Truth in Art, whose publication the *Path* "often reads like a religious tract" (123). Neil Harris (1966) summarizes the dilemma of the nineteenth-century American artist by quoting from a Yale scholar of the era who described the painter's situation as having to work with a "moral alphabet with divine symbols" (303).

After the Civil War, the largest northeastern cities nourished small communities of artists and art associations, but artists largely worked under the precept that art must serve the purposes of moral instruction and be an instrument of virtuous education. Transcendentalists packaged their efforts as a grand syllogism: divinity is to morality as morality is to nature, and as each is to art (Harris 1966). There was no conception of an independent artist or of an art that was autonomous from religion (see Pelles 1963). Green (1966) summarizes in his comprehensive survey of American art that the slow development of American art (relative to the pace of change in Europe) was due to the encumbrance of standards of morality imposed on painters. Without a sustained philosophical tradition that ennobled art within the terms of the Enlightenment tradition, American art was brought under the umbrella of popular religion that emphasized that Christian morality offered the overreaching justification for enterprise, reform, education, and household. Art and literature would serve to justify this grand unity (e.g., Bruchey 1965; Taylor 1976).

To account for developments in the 1880s and 1890s, leading to a full-fledged autonomous art movement in America prior to World War I, requires not only recognition of elite support, but also a framework by which we can interpret changes in the way in which art was defined. This involves two translocations in the institution of art versus nonart: first, a redefinition of art so that it was outside the orbit of religious evaluation, and, second, the clarification that art was hermetic and resided only within a realm of aesthetic — that is to say, interpretive, individualistic, and noncollectivist — values.

Fin-de-Siècle

Art acquired legitimacy at the end of the nineteenth century because of elite sponsorship and the decoupling of aesthetic from religious values, but transformations within art traditions accompanied these changes. Illusionist paintings by Homer Martin, George Inness, and William Harnett became popular in the 1880s, and works by Albert Pinkham Ryder in the 1890s were abstract and sensuous. These developments represented a clear betrayal of the aesthetic canon. One account of this transformation is that the advent of photography pushed artists toward abstraction (Orvell 1989). Other accounts describe the role of international travel, the arrival of immigrant artists in the U.S., and the

growing size of New York's bohemian community. All these may be important, but it can be argued that the transformation had to be endogenous to the institution of art (Schapiro 1978).

By the 1920s the earlier stark contrasts between American art and European art were becoming blurred, and by the 1940s the U.S. emerged as the leading contender in modernism and abstract expressionism. The puzzle is, what intervened between the 1880s and the 1920s that resolved the contradiction between religious and aesthetic values or allowed for the increasing autonomy of art? The explanation lies at the profound level of institutions, but if we are to understand the toggling mechanism, it is useful to consider the shift in public perceptions that accompanied the Arnoldian movement. People's beliefs and values refer to, and are defined by, specific events and charismatic individuals (Weber 1946).

THE UNCOUPLING OF ART AND RELIGION

Matthew Arnold's visits to the U.S. in 1883 and 1886 facilitated the rupture between religion and art, or at the very least defined the emblematic moment of the cleft in the heretofore unified and legitimate institution of culture that embraced sectarian conceptions of morality and art.

Levine (1988) contends that it was Matthew Arnold who "was perhaps the single most important disseminator of such [indifferent] attitudes [about high culture] and [who had] an enormous influence in the United States" (223). At first this appears implausible. Arnold taught poetry at Oxford and was thoroughly High Church, a critic of evangelism, appreciably undemocratic, a proponent of strong state rule, and quite contemptuous of the unkempt and dirty populace (Wilson 1960), Arnold seems an unlikely candidate for weakening the hold that popular religion had on the arts in America; his elitism and antisectarianism would appear to render his polemics unsuited to the climate of opinion in America.

Understanding why Arnold was generally popular and influential is not so straightforward. It cannot be based, as Levine (1988) claims, simply on Arnold's defense of high culture. Americans had already rejected high culture on religious grounds.

Arnold's visits were not merely the occasion for a few wealthy philanthropists to learn more about European art or for intellectuals to justify themselves with an indifferent public. These visits were widely publicized around the nation and his talks received considerable coverage in popular dailies and magazines (Pratt 1988; Raleigh 1957; Wall 1970). Religious tensions were pronounced during these decades, and American intellectuals were wary about overloading secular thought with too many encumbrances, and no American philosopher had ventured to test the line between secular aesthetics and religious morality (Commager 1950). However, Arnold was advantaged in this respect; his allusions to religion in England concealed his considerable distaste for sectarianism from his American audiences.

Arnold was persuasive because his appeal, that "the best that has been thought in the world" is achieved through "the study and pursuit of perfection," signaled to Americans that to be "best" and to pursue "perfection" meant

education and the pursuit of knowledge and was not interpreted as being in conflict with homegrown religion (1897:xx; see Eliot 1932; Livingston 1986; Spender 1963; Trilling 1977; see also Bell 1976). He provided a language to talk about secularism and learning, and an entire industry of criticism flourished in the subsequent decades.

Arnold's accomplishment was to provide a cultural justification for removing art from the dictates of religious morality to those of secular morality. His general popularity and wide influence may be traced to several conditions. First, the American academy was not very theoretical, and neither was Arnold (Abrams 1989). His judgments about the arts were couched in terms of their expressive and redemptive qualities, affirming Weber's observation that art in an intellectualizing period appeals to subjectivity in its competition with religion. Second, Arnold appealed to the pretensions of the upper class and the aspirations of the growing middle class (Lears 1983), while providing an aesthetic framework for urban reform (Bender 1987) and helping academic elites achieve a recognition they had long sought (Cremin 1988; Raleigh 1957).

Third, Arnold spoke the language of "difference" in a milieu that was creating difference. Throughout the 1870s and 1880s there were growing economic inequalities, and accompanying the increase in immigration from central and eastern Europe was a dramatic rise in popular forms of nativism as well as in social Darwinism. The "vogue of Spencer," according to Hofstadter ([1944] 1992), drew on the same language and many of the same principles as the "vogue of Arnold." Specifically, both emphasized that inequalities arising out of competition were inevitable and justified; that ethical hierarchies were more or less coincident with social and economic — if not biological — ones; and both held popular religion and social reform in disdain. It is interesting that one of the most prominent supporters of social Darwinism, Andrew Carnegie, sponsored both of Arnold's visits (Wall 1970) and Herbert Spencer's in 1882 (Hofstadter [1944] 1992).

That Arnold's definition of art had a pretentious note probably helped considerably in the short run: there were growing economic inequalities, a surge of immigrants from eastern and southern Europe, and the urban middle classes were expanding. In subsequent decades, as we shall later see, the art world rejected Arnold's literal interpretations of aesthetic value, and he achieved little lasting fame as an original philosopher (Abrams 1989). Yet the transformations that occurred between 1880 and 1900 make more sense if Arnold is viewed as a pivotal figure than if he is not. It is also useful to show how he advanced different accounts in England and America, because this further helps to illuminate how religion retarded artistic developments in America, although it did not particularly do so in England.

Matthew Arnold in England and America

Culture and Anarchy, which was written in 1869 and revised in 1875 and was subsequently coupled with *Friendship's Garland* in 1899, was intended for an English audience that was steeped in the tradition of the Church of England. Introduced in all editions as an "Exhortation to the Society for Promoting

Christian Knowledge," it is a diatribe against popular sectarianism and a vigorous defense of a higher culture that is both moral and rational. A stark dichotomy is drawn between beauty, intelligence, classical learning, and morality on the one hand, and disestablishment, debasement of authority, superficial learning, and anarchy on the other.

Arnold's (1899) attack on sectarianism is direct: "Because to enable and stir up people to read their Bible and the newspapers, and to get a practical knowledge of their business, does not serve the higher spiritual life of a nation so much as culture, truly conceived, serves." And then he adds, "just what America fails in" (xxvii). Rhetorically, he asks, "how shall we persuade our Nonconformist that his time spent in agitating for the abolition of church-establishments would have been better served in getting worthier ideas of God and the ordering of the world, or his time spent in battling for voluntaryism [sic] in education better spent in learning to value and found a public and national culture?" (201).

Arnold's answer lies in a convoluted argument that places the blame as much on economic liberalism, the idle rich, the lack of state planning, as on the sectarian churches. He defends organized capitalism and contends that the disinterested endeavors of human perfection — which can be traced back to Hellenic times — ought to be protected from the romantic and irrational religious traditions of sectarianism. Arnold (1899) reclaims the arts for classical learning and established religion:

> The great works by which, not only in literature, art, and science generally, but in religion itself, the human spirit has manifested its approaches to totality and to a full, harmonious perfection, and by which it stimulates and helps forward the world's general perfection, come, not from Nonconformists, but from men who either belong to Establishments or have been trained in them. (xvi)

Although Arnold's American lecture tours were widely publicized and well attended (Levine 1988; Wilson 1960), relatively few Americans were actually familiar with his English writings (John 1981; Mott 1938; Raleigh 1957; Shi 1985), and evidently, it was only in the literary salons of Boston and New York that conservative intellectuals actually read and discussed his earlier works (Shi 1985). The lectures given in 1883 and published as *Discourses in America* were based on very different themes from his earlier, English writings. The first discourse, "Numbers," briefly eludes to "excellent Puritan discipline" (Arnold 1885:71), a striking contrast to his description of the Puritans in England as "incomplete and mutilated men" (1899:11). But with this rare exceptional comment about religion, the first "discourse" is largely a discussion about the advantages of America's diversity, while signaling that there may be difficulties with democratic and majoritarian government:

> Perhaps you will say the majority *is*, sometimes good; that its impulses are good generally, and its actuation is good occasionally. Yet, it lacks principle, it lacks persistence; if to-day its good impulses prevail, they succumb tomorrow; sometimes it goes right; but it is very apt to go wrong. (1885:6-7)

Arnold trusted, however, that superior intellectual leaders — the "remnant" — would have the wisdom to lead. The second discourse, "Literature and Science,"

is a defense of the study of literature because it strengthens noble sentiments and profound understanding. In contrast, Arnold (1885) rails against the practical sciences, adding that a "little of mathematics, even, goes a long way" (108).

The final discourse, "Emerson," is a laudatory account of transcendentalism and Emerson, with one mere teasing remark about the sectarians of England: "Carlyle's insistence on labour, righteousness, and veracity pleases [the preachers] of the popular religion; his contempt for happiness pleases them too" (200).

Arnold negotiated the troubled waters in America by emphasizing American exceptionalism (that is, for him, the energy and diversity of the population), praising the homespun tradition of transcendentalism, and appealing to the emerging elite. By stressing the importance of humanistic learning, he gained the support of intellectuals, such as E.L. Godkin, Charles Eliot Norton, William James, and Henry Adams, who themselves had been defensively battling for elite, classical education (see Bender 1987; Shi 1985). His sympathies for strong state control resonated with the regulationists' strategies for moral reform and antivice activity. And, through his friendship with Carnegie, he supported efforts to launch a high arts philanthropic movement that ostensibly transcended religious divisions and that helped to establish bonds between American and European collections and collectors.

Arnold's occasional denunciation of sectarians was aimed not at American Methodists, Disciples, and Baptists, but, duplicitously, at English "dissenters." What Americans learned firsthand from his lectures and also from press coverage consisted of his praise for American industriousness, his high opinion of native literary traditions, and his enthusiasm for the emerging public education system. His defense of high culture was cast in the imagery of social reform, and, most especially, in terms of class (exemplified in the "remnant" and defined in terms of morality and conduct). In was in these establishmentarian and statist terms that he defined culture as "the best that has been known and said in the world" (1897:xx).

The Gordian Knot of Institutions

Increasingly after Arnold's visit, the rhetorical defense of art had less to do with spiritual values and more to do with refined taste, classical canons of beauty, and distinctions of social class. The custodians of high culture that followed the Arnoldian line claimed the moral high ground as composing "the natural aristocracy of virtue and culture" (Shi 1985:157).

The initial steps in the institutional transformation that made possible an autonomous art involved both an elevation of art to a positive good and a connection of refined tastes with status in the social hierarchy. This is the Arnoldian legacy. But this legacy problematized high art because it wedded art to class inequalities and substituted class morality for religious morality. The artist was hardly free. How, then, can we account for the emergence of an American avant-garde in about 1910 (see Shapiro 1976)? One factor was an attack by the Mugwump intellectuals, such as Thorstein Veblen and Edward

Bok, who lashed out both at the American wealthy and at the conservative intellectuals who fit the Arnoldian mold (Shi 1985). To be sure, associations between status and cultural tastes persisted through the twentieth century (Lamont 1992), but these are practical, not ideological, associations, and artists worked after around 1915 as if they did not exist. This claim of autonomy is precisely the rhetorical justification for "art for art's sake" or the avant-garde, much earlier proclaimed in England and the Continent, but never fully accepted or understood in the U.S. in the nineteenth century.

In short, there remained at the end of the 1880s another Gordian knot that had to be loosed before art could acquire an intrinsic freedom that is associated with the avant-garde (Crane 1987), namely, a disavowal of morality, a disregard of class (and patronage), and a denial of tradition. In philosophy and popular thought, social Darwinism was losing its appeal and pragmatism was taking hold. In aesthetic theory the knot was untied by the American philosopher Santayana ([1896] 1955), who in turn was influenced by his contemporary William James, who had repudiated Spencerian evolutionary theory and the doctrines of determinism and inequality. Santayana's accomplishment was to distinguish art, which has no utility, from morals, which involve benefits and have practical significance.

According to Santayana, and consistent with subsequent developments in pragmatism, there are no transcendent criteria that are appropriate for evaluating art and, likewise, no possibility of universal standards or of evolutionary selection, owing to differences among individuals, historical periods, and nations. In fact, art has every right to be "ugly" (Santayana [1896] 1955:29), for its worth lies in its expressive and subjective values. Santayana accomplished two things for the American artists and art. First, he severed art from religious norms, and, second, he situated American art within Kantian aesthetics. Instead of expressing either a manifest historical destiny or the truths of moral and social hierarchies, artworks embodied the artist's creative vision. Lagging behind their European counterparts (Taylor 1976), American artists found by about 1920 a new conception of art, one that was explicitly antibourgeois and anticanonical. In the American context, it was Santayana who provided the secular language for the aesthetic experience. He stated that there is no accounting for art as the "adumbration of divine attributes" (Santayana [1896] 1955:11), and art's only purpose is that it is "*prima facie* and in itself a good" (Santayana 1906:115).

The Armory Show in 1913 was the enacted moment of the birth of modernism in the U.S. Here European avantgardists joined with a small number of American painters. There was a broad representation of early cubism, surrealism, social realism, and early expressionism. And from that moment, art could institute itself as an autonomous dynamistic movement, that is, as a movement that was self-destructive, self-alienating, and progressive (Crane 1987; Hess & Ashbery 1967; Kuh 1965; Shapiro 1978).

The autonomy-of-art position is that "a work of art is the creation of a novel order whose value lies in no end other than its own integrities" (Ross 1982:89). This definition begs the question, autonomy from what? The answers have been autonomy from technology and economy (Benjamin [1936] 1968; Marcuse 1978), from class (Wolff 1983), from politics (Koselleck 1988; Rosenberg 1965), from

theory (Gadamer 1975), and from language (Barthes 1987). From a sociological point of view, art is just a label (Becker 1982). The paradox, of course, is that evaluations become arbitrary in the absence of theory; the impersonal forces of the market of buyers, curators, and critics are difficult to distinguish from their evaluations when the avant-garde rests on the principle of self-destruction. Advanced finance capitalism rests on the conception of arbitrary financial value and an impersonal market. This is not inconsistent with a conception of art that is autonomous, morally agnostic, and independent of any group's claim to a unique identity. Like currency, art became self-referential in its evaluation while universal in its appeal, yet still authorizing subjective escape from capitalist values.

Finally, we can explore the possibility that art is again in the midst of flux, both with respect to the theory-practice problem of meaning (Baxandall 1985; Bürger & Bürger 1992; Danto 1985) and its relation with the economy (Blau 1995; Dubin 1992), but this can be only brief and speculative.

Discussion

Although institutional theory has emphasized the taken-for-granted character and legitimacy of culture, it is argued below that institutional potency depends on antagonisms between institutionalized domains. In the U.S. during most of the nineteenth century, art was so deeply embedded within a religious and moral framework that art could not be justified in its own terms (expressive, as form, texture, decorative). Yet by the end of the century it was possible to forcefully oppose art to nonart as secular aesthetics attained independence from religious morality. Initially, this was achieved within the Arnoldian framework that placed art in the sphere of civilization, positioned at the top of a social and cultural hierarchy.

Santayana provided the philosophical justification for the avant-garde, and he firmly laid down the principles that would free art from every institutional linkage external to the world of art. This extraordinary development pitted art against nonart as a constantly changing but historically self-conscious institutional complex. Nonart included works outside this complex, such as kitsch, popular culture, representational art, and folk art. Art could be abstract "expressionism, photorealism, symbolism, cubism, or whatever," as long as it was poised as if intent on the destruction of what went before. That is, art derived its potency in terms of an aesthetic canon that provided a language for distinguishing art from the commonplace and a vocabulary for describing stylistic transformations (Danto 1981). In an ironic way, art as ideology was independent of economy, politics, and religion, whereas the reception of art depended on its legitimation as a cultural institution within the class structure. The connoisseurship that grew out of the Arnoldian framework can hardly be distinguished from the secular habits of the urban middle and upper middle classes.

What has been recently widely viewed as the death of the avant-garde by cultural theorists (Crimp 1983; Danto 1985; Dziemidok 1985; Jameson 1991; Ventós 1980) has been interpreted by social scientists (Blau 1995; Crane 1992;

Dubin 1992) as the collapse of the distinction between popular and high art. Within the circles of art criticism and theory, art versus nonart now draws on different distinctions: subversive versus nonsubversive (Aronowitz 1994), social praxis versus triviality (Bürger & Bürger 1992), and engagement versus autonomy (Abrams 1989; Bernstein 1992). And in the political arena, art has become the ground of contention and debate both about individual identity — race, gender, and sexuality — and about group rights — those of women, blacks and other minorities, noncitizens, and the politically disenfranchised — and also about the rights and obligations of memorialization.

If high culture has disavowed its own lineage by repudiating its historical roots in the tradition of the "high cultural humanism" of Arnold and Comte (Said 1978:228), can we conclude that it has incorporated popular culture? This line is hard to draw for certain reasons. Both popular and high culture celebrate an autonomy of self and subjectivity, the admixture of codes of the past and the present, and the presentation of the wicked and the playful over the banal and the sublime. It is possible to assert that there are essential differences, assuming that we define high culture as that which is not dependent on commercial support and that we distinguish popular culture that is rooted in the practices of local groups from commercially supported culture.

Commercial elites — the makers of much that passes for popular culture — have *practical reasons* to advance the expressive cultures and traditions of local groups, including the oppressed, whereas the proponents of high culture assert *moral reasons* to advance the expressive cultures and traditions of local groups, especially the oppressed. Commercial elites cannot greatly violate habits of the heart and lodge irony and subversiveness within such habits. Contemporary high culture elites have no such logic, and must — for the sake of consistency — raise questions about their own legitimacy as they raise questions generally about power, oppression, subordination, and group practices. For these reasons, we must view the relations between commercial culture, the high culture tradition of the avant-garde, and versions of local group culture as being more potentially contentious than is commonly recognized in the social sciences, and, indeed, anticipated by Weber, who emphasized the disengagement of modern art.

If Weber failed to anticipate the current reengagement of art with politics and ethics, he provided a sketch of a general framework that can be elaborated to account for major transitions that occur when worldviews are fundamentally contradictory and there are competing vested interests. Yet this suggests a modification of institutional theory, namely, that various potentially contradictory institutional models may exist simultaneously and compete for legitimacy. In large societies, toggled shifts often appear inchoate. Thus, the toggling of institutional change is likely to involve transformations that cannot be predicted with a high degree of precision, yet because competing discourses and cultural practices are contingent on interests, as Weber emphasized, ascendant institutions are amenable to manipulation.

References

Abrams, M.H. 1989. *Doing Things with Texts*. W.W. Norton.

Adorno, Theodor W. 1973. *Negative Dialectics*. Seabury Press.

Archer, Melanie, and Judith R. Blau. 1993. "Class Formation in Nineteenth-Century America." *Annual Review of Sociology* 19:17-41.

Arnold, Matthew. 1885. *Discourses in America*. Macmillan.

____. 1897. *Literature and Dogma*. Smith, Elder.

____. 1899. *Culture and Anarchy, An Essay in Political and Social Criticism and Friendship's Garland, Being the Conversations of the Late Armeinius, Baron von Thunder-ten-Tronckh*. Macmillan.

Aronowitz, Stanley. 1994. *Dead Artists, Live Theories*. Routledge.

Baltzell, E. Digby. 1964. *The Protestant Establishment*. Vintage.

Barth, Gunther. 1980. *City People*. Oxford University Press.

Barthes, Roland. [1966] 1987. *Criticism and Truth*. University of Minnesota Press.

Baxandall, Michael. 1985. *Patterns of Intention*. Yale University Press.

Becker, Howard S. 1982. *Art Worlds*. University of California Press.

Beisel, Nicola. 1993. "Morals versus Art: Censorship, the Politics of Interpretation, and the Victorian Nude." *American Sociological Review* 58:145-62.

Bell, Daniel. 1976. *The Cultural Contradictions of Capitalism*. Basic Books.

Bender, Thomas. 1987. *New York Intellect*. Johns Hopkins University Press.

Benjamin, Walter. [1936] 1968. "Art in the Age of Mechanical Reproduction." Pp. 217-52 in *Illuminations*, edited by Hannah Arendt. Schocken.

Bernstein, J.M. 1992. *The Fate of Art*. Pennsylvania State University Press.

Blau, Judith R. 1989. *The Shape of Culture*. Cambridge University Press.

____. 1991. "The Disjunctive History of U.S. Museums, 1869-1980." *Social Forces* 70:87-106.

____. 1995. "Art Museums." Pp. 87-114 in *Organizations in Industry*, edited by Glenn R. Carroll and Michael T. Hannan. Oxford University Press.

Blau, Judith R., Kent Redding, and Kenneth C. Land. 1993. "Ethnocultural Cleavages and Religious Traditions in the United States, 1860-1930." *Sociological Forum* 8:481-98.

Blumin, Stuart M. 1989. *The Emergence of the Middle Class*. Cambridge University Press.

Bourdieu, Pierre. 1984. *Distinction*. Trans. Richard Nice. Harvard University Press.

____. 1991. *Language and Symbolic Power*. Harvard University Press.

Boyer, Paul. 1978. *Urban Masses and Moral Order in America, 1820-1920*. Harvard University Press.

Bruchey, Stuart. 1965. *The Roots of American Economic Growth*. Harper & Row.

Bürger, Peter, and Christa Bürger. 1992. *The Institutions of Art*. University of Nebraska Press.

Cmiel, Kenneth. 1990. *Democratic Eloquence*. University of California Press.

Commager, Henry Steele. 1950. *The American Mind*. Yale University Press.

Crane, Diana. 1987. *The Transformation of the Avant-Garde*. University of Chicago Press.

____. 1992. "High Culture versus Popular Culture Revisited." Pp. 58-74 in *Cultivating Differences*, edited by Michèle Lamont and Marcel Fournier. University of Chicago Press.

Cremin, Lawrence A. 1988. *American Education*. Harper & Row.

Crimp, Douglas. 1983. "On the Museum's Ruins." Pp. 43-56 in *The Anti-Aesthetic*, edited by Hal Foster. Bay Press.

Cummings, Milton C. 1991. "Government and the Arts." Pp. 31-79 in *Public Money and the Muse*, edited by Stephen Benedict. W.W. Norton.

Danto, Arthur C. 1981. *The Transfiguration of the Commonplace*. Harvard University Press.

____. 1985. "The Philosophical Disenfranchisement of Art." Pp. 11-22 in *The Reasons of Art/L'art a ses raisons*, edited by Peter J. McCormick. University of Ottawa Press.

Demerath, N.J. III, and Rhys H. Williams. 1992. *A Bridging of Faiths*. Princeton University Press.

DiMaggio, Paul. 1982. "Cultural Entrepreneurship in Nineteenth-Century Boston." *Media, Culture and Society* 4:33-50.

____. 1992. "Cultural Boundaries and Structural Change." Pp. 21-58 in *Cultivating Differences*, edited by Michèle Lamont and Marcel Fournier. University of Chicago Press.

DiMaggio, Paul, and Walter W. Powell. 1983. "The Iron Cage Revisited." *American Sociological Review* 48:147-60.

Douglas, Mary. 1986. *How Institutions Think*. Syracuse University Press.

Dubin, Steven C. 1992. *Arresting Images*. Routledge.

Dunlap, William. 1834. *A History of the Rise and Progress of the Arts of Design in the United States*. Vol. 1. G.P. Scott.

Durand, Asher B. 1855. "Letters on Landscape Painting." *Crayon* 1:97-98.

Dziemidok, Bohdan. 1985. "On Aesthetic and Artistic Evaluations of the Work of Art." Pp. 295-306 in *The Reasons of Art/L'Art a ses raisons*, edited by Peter J. McCormick. University of Ottawa Press.

Eagleton, Terry. 1991. *Ideology*. Verso.

Eliot, T.S. 1932. *Selected Essays, 1917-1932*. Faber & Faber.

Gadamer, Hans-Georg. 1975. *Truth and Method*. Crossroad.

Gans, Herbert J. 1974. *Popular Culture and High Culture*. Basic Books.

Gay, Peter. 1986. *The Tender Passion*. Vol. 2, *The Bourgeois Experience, Victoria to Freud*. Oxford University Press.

Green, Samuel M. 1966. *American Art*. Ronald Press.

Hall, Peter Dobkin. 1984. *The Organization of American Culture, 1700-1900*. New York University Press.

Handy, Robert T. 1991. *Undermined Establishment*. Princeton University Press.

Harris, Neil. 1966. *The Artist in American Society*. Braziller.

Herbert, Christopher. 1991. *Culture and Anomie*. University of Chicago Press.

Hess, Thomas B., and John Ashbery (eds.). 1967. *Avant-Garde Art*. Collier-Macmillan.

Hofstadter, Richard. [1944] 1992. *Social Darwinism in American Thought*. Beacon Press.

Holt, Elizabeth Gilmore (ed.). 1979. *The Triumph of Art for the Public*. Doubleday.

____. 1981. *The Art of All Nations, 1850-1873*. Doubleday.

Horowitz, Helen Lefkowitz. 1976. *Culture and the City*. University of Kentucky Press.

Hutchison, William R. 1989. "Protestantism as Establishment." Pp. 3-18 in *Between the Times*, edited by William R. Hutchison. Cambridge University Press.

Jameson, Fredric. 1981. *The Political Unconscious*. Cornell University Press.

____. 1991. *Postmodernism, or the Cultural Logic of Late Capitalism*. Duke University Press.

Jarves, James Jackson. [1864] 1960. *The Art Idea*. Belknap.

Jefferson, Thomas. 1950. *The Papers of Thomas Jefferson*, edited by Julian P. Boyd. Vol. 1. Princeton University Press.

John, Arthur. 1981. *The Best Years of the Century*. University of Illinois Press.

Kasson, John F. 1990. *Rudeness and Civility*. Hill & Wang.

King, William McGuire. 1989. "The Reform Establishment and the Ambiguities of Influence." Pp. 122-40 in *Between the Times*, edited by William R. Hutchison. Cambridge University Press.

Koselleck, Reinhart. 1988 [1959] *Critique and Crisis*. MIT Press.

Kuh, Katharine. 1965. *Break-up: The Core of Modern Art*. New York Graphic Society.

Lamont, Michèle. 1992. *Money, Morals, and Manners*. University of Chicago Press.

Lears, T.J. Jackson. 1983. "From Salvation to Self-Realization." Pp. 1-38 in *The Culture of Consumption*, edited by Richard Wightman Fox and T.J. Jackson Lears. Pantheon.

Levine, Lawrence W. 1988. *Highbrow/Lowbrow*. Harvard University Press.

Lipset, Seymour Martin, and Earl Raab. 1970. *The Politics of Unreason*. Harper & Row.

Livingston, James C. 1986. *Matthew Arnold and Christianity*. University of South Carolina Press.

Lockwood, David. 1992. *Solidarity and Schism*. Clarendon Press.

McCarthy, Kathleen D. 1982. *Noblesse Oblige*. University of Chicago Press.

Madison, James. 1910. *The Writings of James Madison*. Vol. 8, edited by Gaillard Hunt. Putnam.

Marcuse, Herbert. 1978. *The Aesthetic Dimension*. Beacon Press.

May, Henry F. 1976. *The Enlightenment in America*. Oxford University Press.

____. 1983. *Ideas, Faiths, and Feelings*. Oxford University Press.

____. 1991. *The Divided Heart*. Oxford University Press.

Miller, Lillian B. 1966. *Patrons and Patriotism in the Encouragement of the Fine Arts in the United States, 1790-1860*. University of Chicago Press.

Miller, Perry. 1965. *The Life of the Mind in America*. Harcourt Brace Jovanovich.

Minihan, Janet. 1977. *The Nationalization of Culture*. New York University Press.

Morgan, H. Wayne. 1973. *Victorian Culture in America, 1865-1914*. F.E. Peacock.

Mott, Frank Luther. 1938. *A History of American Magazines*. Vol. 3, *1865-1885*. Harvard University Press.

Niebuhr, H. Richard. 1929. *The Social Sources of Denominationalism*. Holt.

Nye, Russel B. 1960. *The Cultural Life of the New Nation, 1776-1830*. Harper.

Orvell, Miles. 1989. *The Real Thing: Imitation and Authenticity in American Culture, 1880-1940*. University of North Carolina Press.

Parsons, Talcott. 1951. *The Social System*. Free Press.

Pelles, Geraldine. 1963. *Art, Artists and Society*. Prentice-Hall.

Perry, Lewis. 1989. *Intellectual Life in America*. University of Chicago Press.

Peterson, Richard A. (ed.). 1976. *The Production of Culture*. Sage.

Pivar, David J. 1973. *Purity Crusade*. Greenwood Press.

Pratt, Linda Ray. 1988. "Matthew Arnold and the Modernist Image." Pp. 81-97 in *Matthew Arnold in His Time and Ours*, edited by Clinton Machann and Forrest D. Burt. University Press of Virginia.

Raleigh, John Henry. 1957. *Matthew Arnold and American Culture*. University of California Press.

Ratzel, Friedrich. [1876] 1988. *Sketches of Urban and Cultural Life in North America*. Trans. and edited by Stewart A. Stehlin. Rutgers University Press.

Rosenberg, Bernard. 1965. "Mass Culture in America." Pp. 3-12 in *Mass Culture*, edited by Bernard Rosenberg and David Manning White. Free Press.

Ross, Stephen David. 1982. *A Theory of Art*. State University of New York Press.

Sadri, Ahmad. 1995. *Max Weber's Sociology of Intellectuals*. Oxford University Press.

Said, Edward. 1978. *Orientalism*. Random House.

____. [1896] 1955. *The Sense of Beauty*. The Modern Library.

____. 1906. *Reason in Art*. Vol. 4, *The Life of Reason*. Charles Scribner's Sons.

Schapiro, Meyer. 1978. *Modern Art: 19th & 20th Centuries*. George Braziller.

Shapiro, Theda. 1976. *Painters and Politics*. Elsevier.

Shi, David E. 1985. *The Simple Life*. Oxford University Press.

Slater, Philip. 1976. *The Pursuit of Loneliness*. Beacon Press.

Spender, Stephen. 1963. *The Struggle of the Modern*. University of California Press.

Swidler, Ann. 1973. "The Concept of Rationality in the Work of Max Weber." *Sociological Inquiry* 43:35-42.

Tarrow, Sidney. 1994. *Power in Movement.* Cambridge University Press.

Taylor, Joshua C. 1976. *America as Art.* Harper & Row.

Timberlake, James H. 1963. *Prohibition and the Progressive Movement.* Harvard University Press.

Tocqueville, Alexis de. [1836] 1945. *Democracy in America.* 2 vols. Knopf.

Toll, Robert C. 1976. *Blacking Up.* Oxford University Press.

Touraine, Alain. 1988. *Return of the Actor.* University of Minnesota Press.

Trachtenberg, Alan. 1982. *The Incorporation of America.* Hill & Wang.

Trilling, Lionel. 1977. *Matthew Arnold.* Harcourt Brace Jovanovich.

Troeltsch, Ernst. 1931. *The Social Teaching of the Christian Churches.* Allen & Unwin.

Trollope, [Mrs.]. 1832. *Domestic Manners of the Americans.* 2 vols. Whittaker, Treacher.

Tyack, David, and Elizabeth Hansot. 1982. *Managers of Virtue.* Basic Books.

Ventós, Rubert Xavier de. 1980. *Heresies of Modern Art.* Columbia University Press.

Wall, Joseph F. 1970. *Andrew Carnegie.* Oxford University Press.

Walters, Ronald G. 1978. *American Reformers, 1815-1860.* Hill & Wang.

Weber, Max. 1920. *Gesammelte Aufsätze zur Religionssoziologie.* Vol 1. J.C.B. Mohr.

____. 1946a. "Religious Rejections of the World and Their Directions." Pp. 323-62 in *From Max Weber: Essays in Sociology,* edited by Hans H. Gerth and C. Wright Mills. Oxford University Press.

____. 1946b. "The Social Psychology of the World Religions." Pp. 267-301 in *From Max Weber: Essays in Sociology,* edited and translated by Hans H. Gerth and C. Wright Mills. Oxford University Press.

____. 1968. *Economy and Society.* 2 vols. Edited by Guenther Roth and Claus Wittich. University of California Press.

____. [1930] 1976. *The Protestant Ethic and the Spirit of Capitalism.* Trans. Talcott Parsons. Allen & Unwin.

Whitehall, Walter Muier. 1970. *Museum of Fine Arts, Boston.* Vol. 1. Belknap Press.

Wiebe, Robert H. 1967. *The Search for Order, 1877-1920.* Hill & Wang.

Wills, Garry. 1990. *Under God.* Simon & Schuster.

Wilson, J. Dover. 1960. "Editor's Introduction." *Culture and Anarchy.* Cambridge University Press.

Wolff, Janet. 1983. *Aesthetics and the Sociology of Art.* Allen & Unwin.

Announcement

Call for Papers

Current World Leaders
International Issues

Current World Leaders, published for 39 years, is accepting papers for publication in 1997.

Each of three international issues will focus on a specific topic::

April: Media and Politics
August: Nationalism and Ethnic Politics
December: Women and Politics

We intend to cover at least tow, and preferably several, perspectives on each topic. For example, the issue on Women and Politics might have one article discussing the changes wrought by women gaining equality and increasing their power within the U.S. Or, it might contain an article presenting a cross-national analysis of the effects on various cultures and political systems of increasing equality for women. We are calling for previously and unpublished papers that present a particular point of view on the topics described above. Articles should be between 4,500 and 10,000 words long.

We also welcome suggestions for other topics.

For more information including deadlines and a style guide, contact:

Thomas S. Garrison
Editorial Director
Current World Leaders
800 Garden St., Suite D.
Santa Barbara, CA 93101, U.S.A.

FAX: 805-965-6071
Phone: 805-965-5010
E-mail: iasb@igc.apc.org

From Protest to Change of Regime: The 4-19 Revolt and The Fall of the Rhee Regime in South Korea*

QUEE-YOUNG KIM, *University of Wyoming*

Abstract

Under what conditions do protest movements develop into revolutionary upheavals and force a change of regime? This article examines the impact of economic deprivation, power difference, public opinion, and critical resources on participation in the waves of a protest movement before and after violent coercion. The empirical setting for our analysis is the historical 4-19 Revolt and the fall of Syngman Rhee s regime in South Korea in 1960. The results indicate that economic deprivation was the most important cause of the collective protest during the first phase of the movement. The authorities' violent response to the protest and the mobilization of public opinion against that violence facilitated protest and transformed it into a major upheaval that overthrew the regime.

Why and how does protest occur, develop into a revolutionary upheaval, and overthrow a regime? The dynamic process of conflict, in which protest becomes revolutionary, involves a sequence of patterned encounters between growing social forces and the power of the state. As Coleman (1990) observes,

> Revolts do not involve a simple two-party structure, with the authorities as one party and the population, considered as a homogeneous mass, as the other. The opponents of the authorities are ordinarily a very small subgroup of the population at the beginning of a revolt, and most of the populace is not active on either side. But even those who are initially uninvolved differ in various ways relevant to the conflict. They cannot be characterized according to a single dimension alone (such as their relation to the means of production, which is how Marxist theory characterizes them). Any population is differentiated in many ways: region, ethnicity, age, class, economic status, and so forth. What is relevant from the view of a theory of revolution is the relation of members of the population to authoritative actions. (490-91)

* *I appreciate the valuable technical comments and constructive suggestions on earlier versions of this article made by Audie Blevins, Burke Grandjean, and David Ashley. This research was supported in part by NSF Grant SES-9115830. Direct correspondence to Quee-Young Kim, Rm. 410, Ross Hall, Department of Sociology, University of Wyoming, Laramie, WY 82071-3293.*

 Social Forces, June 1996, 74(4):1179-1209

Any adequate theory of revolutionary dynamics must consider the relationship between the conflictual encounters and the factors that promote the growth of social forces, to explain how a "small subgroup of the population" gains substantial support for its cause for protest and why the regime suddenly loses the power to control the challenges from below. Tilly (1978, 1993) defines such a situation of extensive political conflict as revolutionary. Specifically, a revolutionary situation combines "(1) the emergence of rival claims to the state; (2) the commitment to those claims by a significant segment of the citizenry; and (3) the incapacity or unwillingness of rulers to suppress the alternative coalition and/or commitment to its claims" (Tilly 1993:10-11).

Protest at first does not usually make an exclusive claim to the state, and thus rarely develops into a revolutionary situation. In the beginning it may appear as legitimate opposition to the authority by a small group of people. Protest becomes illegitimate and thus revolutionary when the authority interprets the demands as a threat and a challenge to its monopoly of power. The conflict between protesters and authorities becomes revolutionary when sovereignty — the state's absolute monopoly on coercive force and the right to exercise that force — is increasingly challenged and rendered ineffective (Tilly 1978). Meanwhile the choices available to challengers and authority "become more and more limited, excluding the options of conciliation and negotiation, while the political conflict becomes more and more unlimited" (Stinchcombe 1965:171). To paraphrase Tilly (1979), the choice of repertoires of repression and contention becomes more and more limited. The rules and procedures of encounter change radically; conflict increases between some members of the population and a changing set of authorities in the nation-state.

Waves of protest can grow rapidly; they may become revolutionary in response to violent coercion as defiance against authority develops into massive demand for institutional change. Claims can escalate into more basic demands, and the basis of support also can expand through waves of protest. Such forces of political upheaval can produce a transfer of power, including a change of leadership and regime.

A revolutionary situation can produce a variety of revolutionary outcomes. "A great revolution entails by definition both a basic split in the polity (a deep revolutionary situation) and a large transfer of power (a serious revolutionary outcome). A revolt may involve an extensive revolutionary situation, but limited transfer of power" (Tilly 1993:15). Some transfers of power occur so gradually, and the actual tenure of power is so brief, that they may not be noticed as a revolutionary outcome. Some transfers of power may occur at the level of regime, not at the level of the state. The major distinction between revolt and revolution is whether the transfer of power occurs in the regime or in the state.[1]

A regime is a formal government with distinctive rules and decision-making procedures; at the same time it is an informal organization in which the center of political power possesses principles and norms (Krasner 1983). "A regime," says Lawson (1993), "determines how and under what conditions the power of the state is exercised" (187). In addition, a regime is a power structure of the state. The state, in turn, is the structure of domination and coordination with the right to monopolize and to exercise coercive force. Unlike a major revolu-

tion, a revolt involves limited transfer of power but it may change the power structure of the state by overthrowing a regime.

A regime becomes weak when the relationship between norms, principles, and rules appears to be increasingly incoherent and contradictory, or when actual practice becomes increasingly inconsistent with the principles and norms. It becomes weaker and may even fall if the challenge to the ruling elites is supported by not only a substantial segment of the citizens but also by some of the ruling elites, and if the authority becomes unable or unwilling to suppress this challenge. Regime change is a change in the principles and norms that govern the rules and procedures.

Authoritarian regimes are more vulnerable than democratic states to protesters demands because charges of contradicting principles may escalate into the issue of leadership and extensive power at the center. The principles and norms of constitutional democracy limit specifically how power is to be exercised. As Lawson (1993) remarks, "by prescribing such limits and establishing the rules of political conduct, constitutionalism underscores the norms of democratic politics and gives rise to a much clearer distinction among state, regime and government. In authoritarian states, however, power is basically undivided and monopolized by the few at the top. No distinctions exist among state, regime and government. The ruling party often identifies itself with the state" (187). Therefore, in such states, the government cannot be changed without changing the regime. A regime change thus resembles a revolution because it implies changes in the dominant power structure of the state.

How then does a revolutionary situation emerge? Specifically, what conditions transform protest into a rebellion? What are the causes and consequences of violent political conflict? Why and how does the protest movement expand so much that the rulers cannot or will not suppress it? The purpose of this article is to identify the conditions that change protest into revolutionary upheaval under the influence of violent coercion.

Despite the extensive literature on protest and rebellion, few studies offer explanations for the transition from protest to rebellion. Even fewer studies explain a national change of regime from a time-series perspective while considering the variation in the local circumstances of collective action. Employing mostly social data and historical information, I combine cross-sectional data on cities with time-series data on collective action to explain the origins of protest and the causes and consequences of violent coercion. The empirical backdrop is a historical revolt in which students and citizens of major cities "rose up" and overthrew Syngman Rhee s regime in South Korea in 1960.

Available theoretical and empirical studies on protest and revolutionary movements mostly come from experiences in Western societies that are different in culture, historical condition and institutional context from those that existed in South Korea in 1960. Although the event that we study belongs to the general category of protest and revolutionary movements, the way in which such processes unfold may vary from culture to culture and the pattern of processes that one finds in one place may not conform to the theoretical models that had been constructed on experiences from another place and time. Often, these theoretical models taken as a whole are not coherent and therefore do not offer any coherent explanations to the pattern of ontologically similar processes.

However, some of the main ideas contained in these models offer powerful insights to organize and illuminate in a theoretical way the story of how people rise up and overthrow their government. Our purpose here is not to apply any dominant theoretical model to the case but to use the main ideas from different theoretical sources to explain the pattern of sequences that involved interactions between forces of opposition and the authority under different local settings. We also identify the causal factors that account for the origin of protest and transition from protest to change of regime. First, I will briefly review the major arguments that purportedly explain the rise of protest movements. I will discuss the conditions of violent political conflict and will relate the effects of violent coercion to the subsequent expansion of protest. Next, I will present a brief historical narrative of the event. Then, combining theoretical insights with the narrative, I will discuss the research design and analytical procedures.

Origins of Protest

One of the most controversial explanations of the rise of the protest movement is economic deprivation. This thesis states that no major protest movement will occur unless a substantial degree of deprivation exists in the population. Deprivation theory emphasizes substantive grievances as a motivation for protest, and predicts that protest will occur because of discontent resulting from economic conditions that impinge on individuals' sense of well-being.[2] According to this theory, unemployment, inequality, poverty, and unsatisfactory living standards are the primary causes of rebellion (Gupta 1977; Gurr 1970; Muller 1972, 1979, 1986; Muller, Dietz & Finkel 1991; Walton 1984; Walton & Ragin 1990). Some deprivation becomes intolerable and fosters revolt, especially when improvements in socioeconomic conditions are followed by apparent deterioration (Davies 1962) or when conditions do not agree with one s expectations in relation to one s capability (Gurr 1970).[3]

Another, equally compelling thesis, based largely on resource mobilization theory, suggests that deprivation does not produce collective action. Poor, displaced, and disenfranchised groups will be "underrepresented in the census of popular activism," according to Aya (1984), because "people with ostensibly the meanest grievances to fight often have the meagerest resources to fight with" (330). Hence the individuals who initiate protest usually are those who are "better situated" (Eckstein 1989:37), a "resourceful part of the critical mass" (Oliver, Marwell & Teixeira 1985) "whose economic basis, social standing, organizational network and political connections afford them the capability to press their claims in the political arena" (Aya 1984:330-31). Mobilization for protest occurs, according to this view, not as a result of the aggregation of individual discontents but because of the ability to use resources that provide the basis for solidarity in organized groups (Aya 1984; Jenkins 1983; Oberschall 1978; Snyder & Tilly 1972; Tilly 1978).

Mobilization for action usually occurs where one finds a large number of ready-made organizations, circles of friends, and especially an organized community of critical consciousness, such as universities and colleges. These places accentuate the existence of such a community, as well as a core of

activists who have more resources than the general population and are more strongly committed to the general principles of justice. Thus students and professionals usually lead the protest because they are "a critical mass of highly interested and resourceful actors" (Oliver & Marwell 1988:1) and command critical resources.

In addition to critical resources, political opportunity structure appears to be crucial in explaining the rise of a protest movement. No protest will occur when the balance of power between the opposition and the regime overwhelmingly favors the few rulers at the top. Protest is common, however, in democratic societies where power tends to be shared through participation in decision-making processes, such as elections. In authoritarian states, power is shared by the people and the government in principle, but in practice it is monopolized by the few at the top. To establish legitimacy, authoritarian rulers conduct elections and seek popular support. The political situation most conducive to collective action occurs when opposition groups are gaining power resources in relation to the ruling coalitions, who meanwhile are losing popular electoral support. The central feature of this, "balance of power" theory is the notion that the differences in power resources over time increase the prospects for both change and for collective action. Decreases in the difference in power resources will increase the probability of protest and political conflict (Kitschelt 1986).

Protest also will occur when protesters expect that public opinion will support their cause. In fact, the mass media in general fertilize the social ground for protest by informing the public about the issues, communicating with potential participants, criticizing the challengers' actions and the policies of authority, and evaluating and dispensing judgment on the political situation. The free press in particular thrives on controversy, exposés, and what Molotch (1979) calls "newsworthiness"; often it criticizes the disjunction between what the authorities ought to do and what they do. The more favorable the existing public opinion appears to be toward the opposition forces, and the more critical of the prevailing authority, the more likely it is that protest will occur.

All these factors — economic deprivation, critical resources, differences in power, and public opinion — can influence not only the origins of protest but also the context of political conflict and the subsequent expansion of protest. Some people may participate first and others later, depending on the political opportunity structures, "which facilitate the development of protest movements in some instances and constrain them in others" (Kitschelt 1986:58).

Causes and Effects of Violent Coercion

When the survival of an authoritarian regime is at stake in the face of the strong possibility of a change in regime, conflict between opposition and authority is likely to escalate into high-stakes revolutionary violence. In electoral years, for example, when a transfer of power is in the offing and there is widespread "dissatisfaction with the current authority system and the newly acquired belief that there is some possibility of change" (Coleman 1990: 484), both the elites and the challengers are likely to resort to violence — the former to maintain the power structure and the latter to change it. People who are most aware of the

contradictions between principle and practice will protest that the rules and procedures for making decisions should conform to principles of democracy. When the authority resorts to violent repression of such demands, the latent structural contradictions become manifest. The protesters learn that one cannot change the rules and procedures of the regime without changing the principles and norms, and the regime itself.

Protest can increase the legitimacy of the protesters cause if their numbers are large. Protesters therefore appeal to the public for support and try to mobilize demonstrations as large as possible. Protest, however, involves a peculiar logic that makes it violent when it is large in scale. Large demonstrations indicate the extent of popular support, and usually are disorganized. The demonstrators are emboldened to express their feelings freely in disruptive and creative ways; they are as much protected as encouraged by the size of the crowd. As DeNardo (1985) observed, "there is safety, as well as power, in numbers" (36). Demonstrations become more and more disruptive as they grow larger. At the same time, the regime becomes less able to control crowds as they grow larger, and the implicit threat of continued disruption provokes violent coercion.

Violent coercion increases the dissidents anger and intensifies their opposition when the regime uses it to impose enduring deprivations on a dissident group (Gurr 1980). Summarizing the main arguments of a comparative study of contemporary revolutionary movements, Gurr and Goldstone (1991) state: "One of the most striking features of our cases [of revolution] is the frequency with which the old regimes followed policies that intensified grievances and focused popular anger on the state. In virtually every instance, threatened governments used force and violence in ways that increased popular resentment and active support for revolutionary movements" (334).

Rather than deterring further protest, violent coercion exacerbates the conflict when the opposition forces are economically deprived and when public opinion toward the regime is generally negative. The economically deprived have nothing to lose because they have little stake in the status quo; they will repay violence many times over if encouraged by public support. Once they mobilize, it becomes easier to do so a second time, and the perceived risks decline rapidly in relation to the increase in moral support from the public. Thus one should expect a transition from protest to a revolutionary upheaval when protest is associated largely with economic deprivation and popular public support, and when the protest encounters violent coercion.

If discontent is widespread and popular support for the regime is weak, and yet the repression is violent, the protest will gain more material support and sympathy from "people who have not suffered directly from the government's excesses" (DeNardo 1985:154). That is, the greater the moral outrage toward violence by the state, the greater the participation by various sectors of society is likely to become. "The more diversified the base of resistance," Eckstein (1989) observed, "the more difficult it is for a state to address the varied grievances of groups concomitantly through force or reform" (49). New participants bring new issues and demands into the protest movement. The movements become a major rebellion not so much because the members are united in a cause as because they are united in opposition to authority. The

original goal of protest develops into more extensive demands, creating an urge to replace not only the policies of the government but also the government itself.

We should expect a direct causal relationship between violent coercion and revolutionary upheaval, especially in authoritarian states. In such states, inherent contradictions exist between the principle of democratic constitutionalism, from which the regime draws its legitimacy, and the decision-making procedures that restrict participation by the opposition. The rules governing the relationship between government and opposition are one-sided in favor of the authoritarian rulers. Some oppositions are tolerated, but they lack the constitutional status of a legitimate alternative government, which the people (in principle) have the power to choose. The resort to force by the authoritarian regime brings to life the contradiction between the rules and procedures of authoritarianism and the principles and norms of popular constitutionalism. At the same time, the use of force by the few at the top implies that the state is unable to meet the challenge within the normal legal-rational framework. Such weakness on the part of the state inspires further insurrection, especially when discontent is widespread and popular support for the regime is weak.

The use of force tends to shift the center of power within the regime from one faction to another, to undermine the regime ultimately from within, and to change the fundamental rules and procedures of the regime. Prime ministers come and go; cabinet ministers are reshuffled; the standard bureaucratic procedures are suspended to meet emergency situations; special legal provisions often are invoked. Debates about proper approaches to conflict divide the decision-making group. In Pareto s ([1901] 1991) words, "The declining elite becomes softer, milder, and more humane and less apt to defend its own power" (59). Humanitarian sentiments then become a platform for rallying the opposition. The tough-minded blame the tender-hearted for hesitating to use force, and force is used. Then the "mild" faction overrides the "tough" faction and blames it for having allowed the revolt to go out of control. Then civilian authority may be transferred to military commanders, who may not want to support the regime and the leaders who have violated the people s trust.

The use of force against popular protest undermines the unity and strength of the regime and sharply reduces the range of options. Either the regime must rule continuously by force, if it can to do, or it will fall. The regime is likely to fall when any further use of force becomes unfeasible and the state no longer can command "the right to exercise control" as a "legitimate authoritative action" (Coleman 1990: 470). In this situation, the authorities cannot or will not use force, and the protest movement succeeds in overthrowing the regime.

The 4-19 (Sa-Il-Gu) Revolt

Power struggles between Rhee and his opponents intensified after the Korean War (1950-53) because Rhee and his party (*Chayudang*) apparently were determined to stay in power at all costs. In 1954 Rhee amended the constitution, over strong opposition in the National Assembly, to allow himself to run for the presidency after his second term, which was expiring. Power struggles between

Rhee and his opponents took a radical turn in the 1956 and 1958 elections, however. In 1956 Rhee won only 56% of the popular vote, compared with the 72% he had received in 1952. The opposition (*Minjudang*) lost the presidential race but won the vice-presidency with 46% of the vote, over Chayudang's 44%.

Two years later, the political scene shifted more dramatically in favor of the opposition. In the 1958 National Assembly election, the ruling party lost five of its seats, falling from a dominating 134 to 129 of a total of 233. The main opposition party increased its representation from 47 to 79 seats, one-third of the total. This number was sufficient to block any major attempts by the ruling party to keep the incumbent president in power through constitutional amendment.

The electoral pattern revealed a sharp geographical division between rural and urban districts. In the 1958 election, for instance, the candidates of the ruling party received 39% of the total vote, predominantly in rural areas; candidates in the main opposition party gained 30% of the total vote, mostly in cities. The opposition established its popular bases in cities and drew its resources from that source in its intensified power struggle against the ruling party. The ruling party, on the other hand, mobilized the government's resources to intimidate the voters into submissive loyalty to the incumbent administration. The power struggle paralyzed parliamentary politics and escalated into the electoral arena, involving the mass media, intellectuals, and students (Q. Kim 1983; Pak 1980).

The Rhee regime stood precariously on a contradictory basis of democratic principles that honored freedom of the press and the electoral system as the basis of legitimacy, on one hand, and repressive authoritarian practices on the other. Most of the press was free to criticize and report irregularities in politics; occasionally, however, the government imposed censorship and even closed newspapers, only to ignite greater dissent and more criticism against the regime.

The number of universities in Korea increased from 28 in 1945 to 85 in 1960. During the same period, the number of college students grew from 7,800 to 142,000. College attendance by youths rose from 4% in 1953 to 12% in 1960, one of the highest in the world (Central Education Research Institute 1967; J. Kim 1985:175-80). Students were recruited on the basis of academic merit from a cross-section of regional and class backgrounds. Education stressed general knowledge rather than narrow skills. Formal socialization emphasized national service as the goal of education, and national issues became the dominant topic of conversation among students. The larger the university population grew, the stronger became the students' concern and interest regarding national politics.

The students' opposition to the existing political arrangements was based primarily on the issue of how democracy should be implemented rather than on the desire to institute an alternative system of government. The students rejected their societies' institutions as backward and feudal in comparison with those of Western countries, and they looked to the Western democracies as successful models of economic development. Their support for democracy was correlated with their own educational experience, which emphasized achievement, competitive standards of merit, and equality of opportunity. These traits frequently conflicted with the traditional, particularistic values of the controlling

elite. The students found, however, that political practices contradicted the highly orthodox principles they had learned at school. They criticized the political authorities for not living up to the ideals of democracy and for not carrying out both the letter and the spirit of the constitution. They perceived the political elites as grossly unprincipled and thus immoral.

The event that galvanized the public and drove students into the streets in the spring of 1960 was a blatant rigging of the election, accompanied by political harassment, widespread and crude repression, and sheer political terrorism. The terrorism was justified largely under the pretext of anticommunism; it was argued that President Rhee must be reelected in order to continue the struggle against Communist North Korea. It was possible for the government and the ruling party to manipulate repression because the state controlled the major institutional resources in the electoral process. Thus they could use these resources to their own advantage as they saw fit.

The rigging was a draconian measure, a high-stakes gamble to defy recent voting trends and public opinion. Major newspapers exposed in detail the scheme to steal the election (copied and compiled from several confidential instructions from the Office of the Minister of the Interior to the local police and election management officers) and campaigned vigorously against it, thus keeping the issue alive. Opposition politicians resorted to various measures such as filibusters and boycotting parliamentary sessions to block the ruling party from winning the election by fraud. Repression and political terror prevailed, however, allowing the election to be rigged.

Opposition party workers were arrested without charges, denied access to selected areas, pushed out of observation posts in voting stations, and lynched by "patriotic" gangs. Police mobilized voters in groups of three to five to vote together in a booth, where everyone was instructed to reveal his or her choices to the others. Prefabricated votes were stuffed into ballot boxes before the boxes were delivered to the voting stations. Some boxes from selected districts were deliberately lost in transit.[4]

To the majority of conscientious students, rigging an election was an outright violation of the rules of fair play. When the authorities themselves violated the very principles that the students had been taught in school, the result was a condition of disjunctive justice (Q. Kim 1991; Moore 1978; Stinchcombe 1965:175), which provided the moral justification for a revolt. When asked to explain the motivation for the "dangerous" action, one student replied, "We were so mad, angry and frustrated by the election rigging and police brutality [that] we wanted to demonstrate to the people and to the government that we were not a living corpse" (S. Kim 1960:96). Another student said, "I was frustrated because the democracy I learned at school was exactly the reverse of what I found in government actions" (97). "We march for democracy and our nation," read a hurriedly prepared placard.

A protest movement developed from the visible gap between what the students' texts described as principles of democracy and what the political authorities practiced under the guise of democracy. Favorable organizational and social conditions prevailed at that time in the students' lives and helped to change their critical consciousness into collective protest. As on campuses in Western countries in the 1960s, Korean students had the tactical freedom and

the ready-made organizations that could be used to mobilize the resources of collective action. The large numbers of students concentrated in university towns were transformed easily into a critical mass for planning and mobilizing demonstrations.

The metropolitan city of Seoul in particular fostered the movement because of its extraordinary number of students. About 80% of the universities, including most of the best, and 60% of all the country's college students were concentrated in that city. It would be comparable to all of Ann Arbor, Berkeley, Cambridge, Madison, and New York were superimposed on Washington, DC. Several elite provincial high schools produced many of the successful students at these leading universities; together they formed dense, hierarchical networks of friendship, and later of leadership, in the movement (K. Kim 1959; B. Oh 1975). The universities were large, with an average of several thousand students at each institution. Often the classes numbered in the hundreds, a situation that made mobilization easy. The position of students in Korean society provided tactical power: they had the institutional resources, the flexible time, and the moral authority to represent the authentic voice of conscience, and their privileged status protected them from harsh punishment by the authorities for their actions.

THE FIRST WAVE OF PROTEST (28 FEBRUARY-19 APRIL)

The protest movement began on 28 February, when hundreds of students from several high schools in Taegu marched into the streets in protest against an unusual directive by the government to attend school on that day, a Sunday. Obviously this directive was intended to discourage youths from showing up at mass rallies held on that day by the major opposition party. No casualties occurred and no arrests were made, but the event inspired further protest movements in other cities. Meanwhile, opposition forces daily resisted the repressive measures employed to "win" the election.

On election day, 15 March, angry citizens of Masan, whose names had been removed mysteriously from the voter registration roster, marched into the city hall asking that voting slips be given to them. Police were called to disperse them. By evening the crowd had grown to several thousand; demonstrators pelted the police with stones and threw back the tear gas canisters that were fired at them. The police brought in fire trucks, but a driver, in a frantic attempt to turn a corner, crashed his truck into an electric pole, triggering a citywide blackout. A jubilant mob battled the police. The police opened fire, killing several youths and injuring many more.[5] The mobs fought back under cover of darkness and ran to the undefended areas of the city, where they destroyed several police stations and government offices.

In Masan the police tortured the arrested individuals indiscriminately, dumped dead bodies into the sea, fabricated evidence, refused to release the identity of those in custody or to state the number of casualties, and framed the whole incident as a communist plot. As a result, fact-finding committees from political parties, major daily newspapers, public interest groups, the National Assembly, and the National Association of Lawyers went to Masan to investigate the incident. The shocking revelations activated public consciousness and

conscience. On 11 April, when a body was discovered in a bay with a police tear gas canister lodged in the left eye socket, citizen unrest in Masan exploded again. As one witness recalled, "Almost every living soul in Masan seemed to have risen to violent hysteria."[6] In battles with the police, 15 persons were killed and 169 were wounded. Several government buildings, police stations, and a notorious politician's house were burned (*Masan Daily*, 17 June, 1960).

After several days of planning, student leaders at the Korea University in Seoul mobilized a major demonstration and led a march of 3,000 students to the National Assembly building. The demonstrators demanded a new, fair election but were answered with few signs of any major political reform. Instead, on their way back to campus, the students were attacked with clubs and chains by hoodlum gangs who later identified themselves as members of the Anti-Communist Youth League, a right-wing organ of the ruling party. On 19 April, the following day, approximately 50,000 students (in a city of 2.5 million persons) fought their way through the police lines and met in front of the principal government buildings in the center of Seoul. These youths were from 30 colleges and high schools, representing almost all of the major universities in Seoul, including Seoul National University.

The government officials were stunned and overwhelmed by the massive upheaval. In the afternoon, police opened fire on the demonstrators in front of the president's office, killing more than 20 persons and injuring many more. This act transformed the protest movement into a major violent rebellion on the day that many people remember as the turning point of the entire episode.

Demonstrations, some peaceful and others violent, occurred spontaneously in Pusan, Taegu, Kwangju, Taejon, and several other cities. In the evening, students and rioters battled the police in a number of cities, killing several policemen and setting fire to major police stations and government buildings. In Seoul and elsewhere, 186 persons were killed and about 1,600 were wounded during the day and night of 19 April. Martial law was declared; on the following day, soldiers took control of the streets of Seoul and four other cities.

THE SECOND WAVE OF PROTEST (20 APRIL-28 APRIL)

For a few days a deceptive calm prevailed in the country; the politicians of the ruling party interpreted this as a sign of impending victory. Yet the use of force on 19 April — indiscriminate and uncontrolled — had seriously undermined the people's trust and transformed the conflict into a revolutionary situation. This new situation made it difficult for the Martial Law Command to use force again. The government s inability to control the massive protest movement by any means other than force shifted the power from the civilians to the military.

Then a leading politician of the ruling party — Vice-President-"elect" Lee Ki-Pung — made the unfortunate announcement that he was "considering resigning." This move was perceived widely as an arrogant manipulation. The military could not and would not support the civilian authority. The top-ranking martial law commanders feared that by successfully controlling the rebellion, they would end up serving the narrow, corrupt interests of the ruling party rather than the national interests. They believed that controlling the rebellion would be possible only by using massive force and probably causing

a civil war. Such an option was viewed as unfeasible and unfathomable in a country that had been devastated by an inconclusive civil war in the recent past.

International pressures, especially from the U.S., intensified during this phase. The U.S. government refused to support repressive measures against protest; this refusal dramatically undermined the legitimacy and effectiveness of the government. The Korean Army, which was subject to the American Military Command in its chain of operational authority, could not provide the regime with the loyalty and logistic support necessary to suppress the forthcoming upheaval.

On 26 April, defying a ban on demonstrations, a group of about 100 university professors in Seoul broke the silence and marched into the streets, loudly demanding Syngman Rhee's resignation. This was one of the first demonstrations of its kind during the entire affair. In the evening, several hundred thousand persons rioted, ransacked the headquarters of the ruling Liberal Party (*Chayudang*), and burned buildings and houses that belonged to notorious politicians. Though brief in duration, the protest was a revolutionary upheaval in spirit and purpose because the intent was to overthrow the regime. Rumor had it that many more hundred thousand persons were planning a major upheaval on the following day. President Rhee resigned the following morning, 27 April, in a desperate move to prevent further bloodshed while a throng of people moved into the streets.[7]

The event is called by various names: the April Revolution (K. Lee 1984), the 4-19 Righteous Uprising, the 4-19 Revolt (Q. Kim 1983), or simply the *Sa-Il-Gu* ("April 19th"). It gave rise to a potent legend of righteous students fighting political corruption as a touchstone of revolutionary movements, and inspired subsequent student radicals to repeat that extraordinary success. The 4-19 Revolt was revolutionary in that it changed in a sudden and violent way the rules, procedures, norms, and principles of the regime, but it was not a revolution. Like the several rebellions that Walton (1984) studied, this revolutionary situation was fully developed but did not progress to the point of coalescing into a unified "alternative polity." The situation fell short of revolution in that the protesters did not claim to be the government.

Unlike the patterns found in classical social revolutions, the revolutionary outcome of the 4-19 Revolt was limited to changes of leadership, policies, and government in the regime without changes in the state and in property relations. The revolt involved mainly urban areas, leaving most of the rural population unaffected. The urban participants, however, included not only students but also a substantial portion of the lower classes, especially the unemployed. Almost half of the 186 casualties, for instance, were college and high school students. The next largest category was the unemployed, constituting about 25%. About 20% belonged to the working class, including day laborers, factory workers, mechanics, and taxi drivers. Several others were listed as clerks, officials, and staff members of political parties. A few were listed as merchants and farmers (*Sa-Il-Gu* Yujokhoe 1971). In Masan, 75% of the wounded were listed as unemployed (*Masan Daily* 1960, 3).

The scale of this revolt is comparable to that of any major revolutionary upheaval in any country. In Seoul alone, the revolt involved almost 100 organizations with 121,850 participants, amounting to 5% of the city's total

population or 18% of the economically active population. This proportion is similar to the level of workers' participation in revolutionary activities during the periods of revolt in France during the 1830s and 1850s (Tilly 1974, 1986, 1989).

In Korean political culture, the 4-19 Revolt is glorified for evincing the spirit of rebellion against injustice and the courage to seek democratic rights and liberties. At the same time, however, it is criticized for a lack of creative vision and for the political disorder it produced. Nonetheless, it remains as a touchstone of Korea s struggle for democracy.

The Process Model

The 4-19 Revolt was a convergence of patterns of protest into a revolt against the regime in what Oberschall (1980) characterized as a "loosely structured collective conflict": "Hundreds of groups and organizations — many of them short-lived, spatially scattered and lacking direct communication, a single organization and a common leadership — episodically [took] part in many different kinds of local collective action" (45). Largely in interaction with authority, such a loosely structured collective protest was transformed into a rebellion: the first wave of protest, followed by violent confrontation between protesters and authority, became a revolutionary upheaval in the second wave. Radical demands were made for the resignation of the President and the end of the regime. A significant segment of the citizens supported these demands, which the rulers and the military became unwilling and unable to suppress.

When we regard conflict as a dynamic process unfolding in a series of stages, our attention is directed to transitions between stages. In particular we may ask why protest escalates to the point where violent coercion comes into play, and why violent coercion increases and raises the level of protest to rebellion. Although the escalation process may be complex and although a full account of its dynamics is beyond the scope of this study, one still can make causal inferences from an analysis based on a process model that distinguishes each of these phenomena conceptually and relates them to one another in a series of interconnected equations.

The process model combines the notion of events and actions as temporally ordered, sequential phenomena with theoretical concepts that organize and transform temporally conditioned factual disorder into a timeless placeless causal structure. This model is based on two major premises. First, causes and consequences are partly contingent on the order of an action or event in an entire sequence of temporally related patterned processes: a single action or event may have different antecedents and effects if it occurs at a different point in one sequence than in another. Actions and events, however, are not discrete, disparate or meaningless pieces of history; they are bound together into a coherent and configured conceptual whole that gives meaning to each of its elements.

Second, the heterogeneous structural conditions to which individuals and groups belong at any time and place influence the probability of widespread participation in collective action. Construction of a process model begins with

identifying a colligated pattern of events and actions in a historical context, translating this pattern into concepts, and arranging the concepts into a time series and a verifiable theoretical form. In short, the model is formulated by interpreting the temporal and spatial relatedness of a historical phenomenon.

This strategy is suitable for a case study of revolt, in which one seeks generalizations about the dynamic process by combining the cross-sectional spatial comparative analysis of variations of local circumstances within a nation with the temporal sequence of events and actions. I analyze the first phase in terms of the theories of the origins of protest, and the second from the perspective of the consequences of violent coercion. I suggest that change of regime is the result of the combined effects of the following three linear regression models, Y_{t1}, Y_{t2}, and Y_{t3}:

$$Y_{t1} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon_1$$
$$Y_{t2} = \beta_0 + \beta_1 X_3 + \beta_2 X_4 + \beta_3 Y_{t1} + \epsilon_2$$
$$Y_{t3} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 Y_{t2} + \epsilon_3$$

where Y_{t1} is *protest* at time 1 (*protest* 1), β_0 is the intercept, β_1, β_2, β_3, and β_4 are parameters to be estimated for the respective variables, X_1 is *economic deprivation*, X_2 is *public opinion*, X_3 is *power difference*, X_4 is *critical resources*, Y_{t2} is *violent coercion*, Yt_3 is *protest* at time 2 (*protest* 2), and ε_1, ε_2, and ε_3 are unknown error terms. I compare variable scores across the units of analysis, 23 cities, for different types of collective action within the recognized periods. The variables are urban characteristics and activities occurring at the city level. Each equation represents a panel for each distinct phase of the process as ordered sequentially.[8]

The main feature of the process model is a series of regression equations in which the dependent variable in one equation becomes one of the independent variables in the next. The dependent-turned-independent variable provides a crucial theoretical link between equations. Such a design, however, may create multicollinearity when the independent variables are highly correlated and when two or more of these variables that show high correlations with the dependent variable are used as independent variables in the next equation. In such a case, the model must be refitted with the data. To minimize collinearity, I opted (among various strategies) to exclude from the subsequent equation the variable that appeared to have the strongest influence on the dependent variable in the earlier equation.

In addition, a process model may suffer from autocorrelation if the dependent variable is time-series data from the same flow of events but is measured separately at two or more different time periods for conceptual and analytical purposes. Earlier protests, for instance, will be correlated with later protests. Note that Y_{t1} (*protest* 1) and Y_{t3} (*protest* 2) are the same protest actions in a continuous flow of events. Protest may generate protest, and the pattern of waves may reveal some particular mechanism that describes the evolution through time. We assume that this may be the case, but the more interesting issue is the extent to which different exogenous factors contribute to the variation of patterns from one stage to the next. Therefore I separate the first phase of protest from the second phase and exclude from equation 3 the impact

of earlier protest. This strategy minimizes autocorrelation and improves the fit between the model and the data.

Protest 1 (Y_{t1}) and *protest 2* (Y_{t3}) occurred at different points in time in relation to the violent coercion, and thus should have different causes and consequences. Each city responds differently not only to the different local characteristics and opportunities but also to the threats and responses by authorities at local and national centers of power. Y_{t1} and Y_{t3} are collective actions; they are different from Y_{t2}, which is violent coercion committed mostly by the agents of the regime. Thus Y_{t2} should have somewhat different causes and consequences. Violent coercion may be a response to protest; this, in turn, produces collective violence as a by-product of groups' capacity for collective action (Snyder 1976; Snyder & Tilly 1972), not particularly as a reaction to public opinion. Violent coercion, however, may either constrain or expand protest movements, depending on public opinion about it. In other words, public opinion may not be directly relevant to explaining the causes of violent coercion, although it may be crucial for explaining the consequences. *Protest 1* is a crucial linkage variable; thus, in equation 2, it is used as one of the independent variables.

In a preliminary analysis, I learned that some variables show high correlations with *protest 1;* accordingly, I excluded one of these variables from equation 2. Equation 3 considers the impact of violent coercion on the second wave of protest in conjunction with the presumed effects of the factors that produced the first wave, but to avoid simultaneity (*protest 2* follows *protest 1* closely), I did not include the effect of the first wave of protest on the second wave. If I had done so, the regression model must have suffered from autocorrelation.

Variables and Measures

Protest is public opposition to the rules and decision-making procedures of a regime, in which objections are expressed openly and collectively. *Protest 1* and *protest 2* are the natural logarithms of the nominal value of the total number of demonstrators per 1,000 economically active persons in the population. Most demonstrations took two or more hours in the course of a day. Unlike strikes, the organizing agency — in the form of a social movement organization — existed only for the duration of the demonstration, although many of these groups emerged from existing schools and represented those institutions. Many demonstrations occurred spontaneously in the streets, largely precipitated by rumors or incidents.

For each city, I summed the number of demonstrators per 1,000 members of the economically active age group for each of the two periods (from 28 February to 19 April and from 20 April to 28 April) and transformed it into logs. Any city during any day in the first period and the second period had an equal distribution of probability of collective action. The pattern of authority, the institutional means of control, and the dominant leadership changed from one period to the other; so did the relationship between coercion and the protest movement. During the first period, the government used violent repression through

national and local police. In this period, it was primarily the national police that tried to control the series of massive demonstrations.

The purpose of measuring the dependent variable in two different contexts is not only to identify the structural origins of protests and to distinguish them from developmental dynamics, but also to estimate the differential structural effects on the distribution of protests, controlling for the effects of violent coercion. This procedure permits more time-oriented causal inference of the effects of violent coercion on the second wave of protest and of the effects of conducive factors in each of the cities on the process of transformation from protest to regime change.

As a measure of *economic deprivation,* I use the unemployment rate as a percentage of the active labor force aged 15-59, as reported by the census. The census defines as unemployed those who are "not gainfully employed" in any occupational sectors including industrial, commercial, and professional fields. Because no unemployment compensation system existed in any form, the index reflects not only the extent of economic deprivation but also the general deterioration of the standard of living across the main economic sectors of urban life. The unemployed do not have organizational resources, and tend to blame themselves rather than others for their own misfortunes. Yet because they have very little stake in the status quo and ample time on their hands, they may cooperate spontaneously and may easily become rebellious when the opportunity of change arises.

Power difference is measured in terms of the difference between the percentage of opposition votes and the percentage of ruling-party votes per city in the 1956 and 1958 elections. The index reflects the extent of opposition to the regime as the degree of support for opposition candidates over the candidates of the ruling party increased during the two national elections preceding 1960.[9]

I measure distribution of *public opinion* by indexing newspaper readership of a certain political orientation. The relationship between newspaper readership and political orientation is reciprocal. The newspapers compete with one another to increase influence and revenues by increasing subscriptions and acquiring advertisers. They report events and editorialize on issues as their imaginary readers would like them to be reported and evaluated. In turn, the readers who already are critical of the government may subscribe to a newspaper that takes positions similar to their own. The two elements rarely match perfectly, but a rough correspondence exists between the sentiments of the public and the circulation of a certain kind of newspaper, whether the paper is critical of the government or favorable toward the opposition. Thus the decision to subscribe to a particular newspaper may represent reinforcement as well as reflection of political predispositions. In this way the leading newspapers cultivate a network of public opinion and foster common attitudes toward prevailing issues. They provide a basis for coherence and unity, among widely scattered potential participants, in collective action and effective mobilization for massive common action.

The press in Korea has a long history as the champion and popular advocate of national and public causes rather than solely as a source of entertainment (Jacobs 1985:53-54). "At least 60 percent of the 42 daily newspapers," observed Robert Oliver (1957), "are anti-government in tone and together

they probably comprise at least 80 percent of the total circulation" (85). Contradictory as it may sound, the freedom of the press to criticize the government existed despite repression by the regime. The major dailies, all concentrated in Seoul, commanded approximately 62% of the total circulation throughout the nation and were decidedly antigovernment. From 1945 to 1960, for instance, the leading opposition newspaper, *Dong-A Ilbo,* devoted 191 of a sample of 401 editorials to evaluating the government's policies. About 60% of these editorials were highly critical (In 1970). A content analysis of another leading opposition newspaper, *Chosun Ilbo,* showed that one of every three editorials published between 1954 and 1960 criticized the government's policies and leaders for violating democratic principles (Cho 1969). *Kyunghyang Shinmun,* until the government suspended its publication in 1959, carried an editorial every other day "to promote a fair election," and criticized the repression and "dictatorship" of the ruling regime. The *Hankuk Ilbo* often printed front-page editorials on issues related to the "democratization of Korean politics." The circulation of the four major dailies reached an estimated total of 500,000, and one of every three households in the major cities subscribed regularly to one of these papers (Ministry of Public Information 1961).

The index of newspaper readership is measured by the percentage of households per city that subscribed to one of the four major dailies in 1960. (The exact month of the survey is not known, but presumably it was conducted in early 1960 by the Ministry of Public Information. The results were kept confidential for several years.) Newspaper readership during this period of Korean history indicates the extent of public opinion critical toward the regime. Unfortunately, no public opinion surveys were held in any comparable form across cities, and we have no day-by-day information from polls about the events during this period. Therefore I use newspaper readership as a proxy for public opinion critical toward the regime during the period under study.

Critical resources are the availability and distribution of manpower necessary for mobilizing collective action. Usually such resources exist in various forms and degrees of organization, formal and informal networks, existing voluntary associations, or groups within institutionalized organizations. The term *critical* performs double duty: it refers to the intellectual preoccupation of a segment of the population with distinguishing right from wrong; at the same time, it denotes a threshold effect of the existence of such a population in mobilizing for collective action. The measure of critical resources is the population of high school and university students as a percentage of all persons aged 16-24. The major premise is that protest is likely to occur and grow where the probability of association is high (Oberschall 1973), and that the probability of collective action will increase as the size of the intellectually critical community increases. Intellectually, students tend to be critical toward the contradiction between the principles they learn in the classroom and the practices they observe in politics. Physically, schools provide ready-made organizational resources for mobilization. The size of the student population in a city affects the probability of protest both directly and indirectly because someone is likely to initiate the movement if many people are present. The decision to initiate protest depends on the expectation that more people will be available to support the cause. At

the same time, the decision to join a movement may depend on the numbers of persons already participating (Granovetter 1978).

Violent coercion includes repressive measures by the government, such as arresting, wounding, and killing demonstrators. It is "the cost of collective action to the contender resulting from interaction with other groups; as a process, any action by another group which raises the contender's cost of collective action" (Tilly 1978:55). In a revolutionary situation where the political language becomes violent, two faces of violence appear: civil violence, composed mostly of property damage, injuries, and deaths, and violent coercion. Violent coercion excludes property damage, injuries, and deaths of agents of authority. The use of such coercion "signals an inability to resolve conflict politically through channels that are regarded as legitimate" (Jackman 1993:158).

The Rhee regime intensified repression of political opposition as the national election approached. As early as 1959, it suspended publication of a major newspaper (*Kyunghyang Shinmun*). Several campaign workers were arrested on nebulous charges in the early months of 1960. During this period, as Scalapino (1961) observed, "the opposition Democratic Party is being intimidated, its leaders threatened, its press restricted, and its general rights violated. The charge of 'treason' is being used loosely to justify these actions" (509). The repression became violent as protests against the electoral fraud spread to several cities, beginning in early March 1960. The police were armed; they opened fire on demonstrators in Masan on 15 March, election day, and later in other cities as well.

I measure violent coercion in two steps. First I take the common logarithm of the number of protesters per city arrested and wounded during demonstrations by the agents of authority between 28 February and 19 April, 1960. The government agents did most of the wounding and killing. The occurrence of deaths in civil strife has a special meaning and provokes strong emotions because the authority, which has an absolute monopoly on force, has violated the trust that it will use its power properly. Therefore, as the second step, I add the actual number of people killed by the government's forces during that period to the logs of the number of persons arrested and wounded per city during the same period. Such a procedure is justifiable because the use of physical force is a crucial index of the survivability of the regime as a legitimate system. Ironic as it may sound, the more frequently the regime resorts to force, the weaker it becomes. The availability of force increases power, but the use of force against protesters undermines that power, especially when the public views the protest as legitimate. I gathered the information from the records of the casualties, collected by the Association of Bereaved Families of the 4-19.

The violent repression that occurred during the first period (28 February-19 April) was reactive coercion. It provoked anxiety and flight in some persons but created anger and aggression in many others. In the second period (20 April-28 April) the government relied on the military for law and order, but the cumulative effects of the use of force (in the first period) and the reluctance of the military to use force (in the second period) created opportunities for a revolutionary upheaval. (Appendix A lists the measures of dependent and independent variables and cites the statistical sources of the data.)

Analysis and Findings

The pattern of protest occurrences and the extent of violent coercion are reported in Figures 1 and 2. The protests occurred in clusters: first around election day against electoral fraud, and then in reaction to the brutal police repression of protesters. A high, positive correlation exists between the scale of mobilization and the magnitude of violent coercion. Violent coercion increased as demonstrations became massive, especially in Masan and other cities where police opened fire on protesters. Citizens rebelled in several cities, especially Masan, where the confrontation was most violent. On 19 April, both demonstrations and casualties reached their highest levels. Violent clashes had a greater impact on the crisis than did nonviolent demonstrations in that the deaths of protesters deepened the meaning of the protest and placed the responsibility for the crisis directly on the nation's leaders. The last phase of the revolt witnessed both a massive upheaval and a large number of casualties across major cities in South Korea.

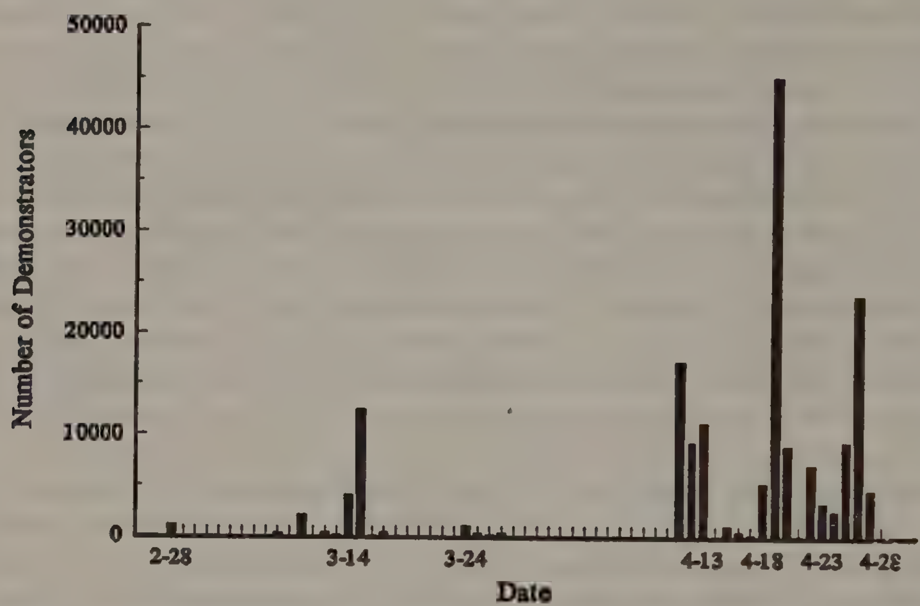
We now assess the effects of indigenous social structure on the origin of the revolt, while considering the effects of earlier protests and violent coercion on later protests. Table 1 shows the ordinary least squares (OLS) unstandardized regression coefficients of the independent variables on *protest 1*, violent coercion, and *protest* 2. (Appendix B presents a correlation matrix of the variables.)

The regression analysis of the *protest 1* model reveals somewhat surprising findings. Contrary to my original expectation, the effect of the history of political contention, as measured by differences between votes for the opposition and for the ruling party, appears to be weak. So does the effect of student population. Also surprising is the negative, weak effect of newspaper readership on the initial stage of the revolt. As displayed in Table 1, column 1, none of the regression coefficients are statistically significant except the relationship between unemployment rate and *protest 1*. A 5% difference in the unemployment rate, for instance, is associated with almost a full unit change in the log of the magnitude of protest, from 100 to 1,000 and from 1,000 to 10,000 on average. It is evident that some opposition vote and a student population are necessary if an initial protest is to occur, but economic deprivation appears to be the preponderant influence.

The finding that economic deprivation alone affects the probability of protest is surprising because it is not compatible with the central argument of the widely accepted resource mobilization perspective. This perspective assumes that the level of deprivation is fairly constant over time and therefore that deprivation would have no significant effect on collective action. But my findings, however, contradict resource mobilization theory and support discontent theory. It appears that extensive unemployment concentrated in urban areas can produce spontaneous cooperation and collective action in the streets. Presumably this reaction is precipitated by some incidents; also and more importantly, it is inspired by political opportunities.

The finding that some opposition vote and a student population, in conjunction with extensive unemployment, may be necessary conditions for protest suggests that discontent can lead to collective action when the balance of power between the opposition and the ruling party changes in favor of the

FIGURE 1: Protest Mobilization in South Korea, 1960



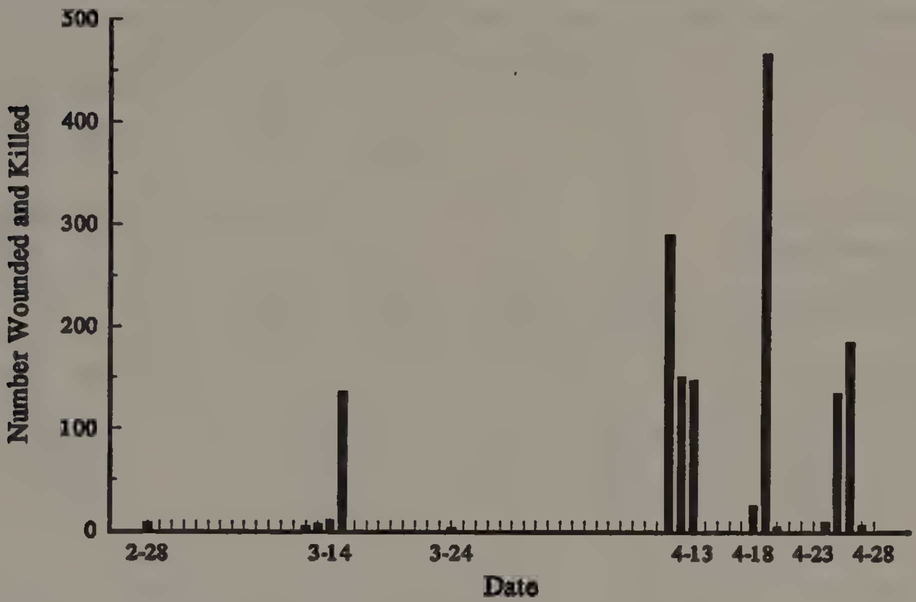
opposition in such a way as to create political opportunities and encourage collective action on the part of the deprived, and when the student population is large enough to be able to launch similar struggles against the regime.

The fit between the model and the data is good enough to create confidence in the robustness of the findings. Regression diagnostics, as indicated by the variance inflation factor ($\sqrt{VIF}$), show no signs of collinearity. The Durbin-Watson *D* statistic shows that the model suffers from no serious autocorrelation. The adjusted R^2 value indicates that the independent variables considered here are important factors in protest. The four independent variables together explain 42% of the variance in the dependent variable during the first wave (adjusted $R^2 = .420$).

The causes and consequences of violent repression are important in understanding the transformation process of protest. The larger the protests, the more likely that they will provoke violent repression. The findings reported in Table 1, column 3 indicate that *protest 1* is the most significant influence on violent coercion. In addition to the size of the protest, student population seems to be crucial in provoking violent repression. Regression diagnostics indicate that none of the independent variables suffer from multicollinearity. The Durbin-Watson *D* statistic indicates that autocorrelation does not exist, thus providing the basis for confidence in the reliability of the estimates.

About 62% of the variance is explained by the three variables included in the model. A large concentration of students (usually in large cities) produced large-scale protests, which escalated into violent confrontations with the police. Two major social groups — the unemployed and the students — seem to have produced different outcomes: the unemployment rate made the difference in the

FIGURE 2: Violent Coercion in South Korea, 1960



overall size of the protest movement, whereas the student population seems to have been decisive in eliciting violent coercion. The violent confrontation between students and the authorities elevated the meaning of protest from local concerns about deprivation to national issues of democracy and justice. Violent repression of students in particular destroyed the legitimacy of the use of force, mobilized public opinion, and galvanized national opposition to the regime.

Protest that occurs in the context of political conflict at the national level is likely to trigger further protests. Protest generates protest and becomes a revolutionary upheaval as the interaction between protesters and the authorities grows violent. During the first period of crisis (28 February-19 April), the government used intermittent but violent repressive measures on students and other citizens in Taegu, Masan, Seoul, and several other cities. Such inconsistency of coercion tends to inspire more protest in some cases and less in others. Protests occurred in the second period (20-28 April) in cities where they had been staged in the first period. These differences in the pace and magnitude of protest may be attributable in part to the effectiveness or, ironically, the ineffectiveness of coercion. The factor that makes coercion most ineffective and thus inspires more protest seems to be the combination of extensive violence and public opinion critical toward the regime.

One of the most striking findings of our study is that violent coercion changes protest movements into revolutionary movements. As reported in Table 1, column 5, only newspaper readership and violent coercion appear to be statistically significant at .01 level. The effects of unemployment rate and opposition vote on *protest* 2 are not significant. These selected variables together explain approximately 65% of the variance in *protest* 2, revolutionary upheaval. The model seems to fit the data well. Regression diagnostics reveal no indications of autocorrelation or multicollinearity; thus the estimators appear reliable.

TABLE 1: Ordinary Least Squares Regression Analysis for Protest 1, Violent Coercion, and Protest 2

	Protest 1		Violent Coercion		Protest 2	
	Coef.	$\sqrt{VIF}$	Coef.	$\sqrt{VIF}$	Coef.	$\sqrt{VIF}$
Unemployment rate	.1960**	1.095	—[a]		.0331	1.185
Newspaper readership	-.0122	1.322	—[a]		.1220**	1.173
Opposition vote	.0107	1.161	-.0212	1.156	-.0021	1.166
Student population	.1662	1.267	2.6414*	1.156	—[a]	
Protest 1			7.9657**	1.208	—[a]	
Violent coercion					.0373**	1.230
Constant	-2.9576		-61.5470		-0.1213	
Adjusted R^2	.420		.623		.651	
Durbin-Watson D	1.869		1.690		1.229	
(N = 23)						

[a] The variable was not entered.

* $p < .05$ ** $p < .01$ (two-tailed test)

The findings suggest that transition from protest to revolutionary upheaval is spurred by mobilization of public opinion critical toward violent coercion. Violent coercion tends to escalate protest in cities where public opinion critical toward the regime already exists to a substantial degree. In other words, public opinion seems to have as much impact on the transition from protest to revolutionary upheaval as does violent coercion.

Newspaper readership seems to transform protest into revolutionary upheaval by politicizing, moralizing, and condemning violent confrontation between the authorities and the protesters. By simply reporting the news, newspapers tend to enlarge the network of potential participants; by evaluating what the government did or should not have done, they draw the battle lines; and they elevate local concerns into national issues, unify protest movements with similar purposes, and characterize the movement in a way that is more meaningful and larger than life. By taking a broad national perspective, the newspapers expose the contradictions between democratic constitutionalism and authoritarian practices and condemn the narrow political interests of the regime, against which very few elites can find a justifiable defense.

Violent coercion seems to have exactly the opposite effect to what was intended, namely increasing the power of the regime. It dramatically exposes the authoritarian power behind the thin veil of democracy and undermines the moral legitimacy of the regime. More important, violent coercion seems to divide the elites from within — separating the "mild" from the "tough" faction — and thus makes it difficult for the regime to act effectively. Therefore, ineffective and indiscriminate coercion provokes revolutionary upheaval. The government's resorting to indiscriminate violence destroys the system of legal-

rational authority and provides the opposition with a license to commit violence; it invites eruption not only by the unemployed or students but also by substantial portions of other segments of the urban population. At this point the protest becomes a revolutionary upheaval because the issues of discontent, violence, and law and order are now politically and ideologically charged as the result of misrule.

Discussion and Conclusion

This study provides substantial evidence that protest begins with economic deprivation and is transformed into revolutionary upheaval not only because of violent coercion but also because of generally critical public opinion regarding the government. Regression analysis of data that rely on the characteristics of local social structures rather than on national stimuli supports the thesis that a regime changes when economic deprivation is followed by violent coercion and by mobilization of critical public opinion.

We must recognize the limits of regression analysis and of the ability to draw conclusions about the dynamics of revolt at the national level by analyzing variations in local conditions. Our findings must be evaluated in terms of the context of the time and the national setting. In understanding the connection between economic deprivation and protest, we must take into consideration the national atmosphere in South Korea in 1960; This atmosphere was charged politically by the overall climate of prospects for change, by the approaching elections, and by the widely publicized fear of electoral fraud. When we consider these national circumstances, it is not hard to understand the effect of economic deprivation on protest.

To be sure, as is commonly observed, the unemployed may suffer a high level of deprivation, but they have little propensity to protest. Although they have time on their hands and resent their situation acutely, these "resources" do not easily produce an organized revolt because of a "pathology" that paralyzes them and prevents them from acting together (Horwitz 1984). The economically deprived blame themselves for their misfortune, isolate themselves from the rest of the society, are unwilling to treat private difficulties as a public issue, and are ambivalent toward authority. Pervasive unemployment in major cities, however, is more likely to transform private misfortunes into a public issue and also to justify protest when the political opportunities change.

The national electoral sentiment in the late 1950s was turning decisively against the regime. While the rural areas remained under the influence of the ruling party, the cities became the bastion of opposition. Eighteen of 23 cities voted consistently for the opposition candidates during the 1956 and 1958 elections; only five showed some support for the government candidates. No direct connection exists between support for the opposition and the subsequent protest movement, but it was in this political climate — increasing opposition to the regime — that widespread grievances found political expression. In short, economic deprivation becomes a factor in the emergence of protest when it is so extensive and so highly politicized that individual miseries become a public, political issue for which the existing government is viewed as responsible.

In addition, support for the opposition party at the local level is a weak predictor of protest. It shows only negligible influence on the initial stage of revolt and exerts no direct influence on the revolutionary phase of the movement when the other independent variables are held constant. This finding seems to reflect the fact that the opposition parties were not strongly in opposition. These parties were defending their own interests and those of a few political groups rather than broad class interests. The main opposition party, *Minjudang*, was conservative in its policies and deliberately avoided identifying with the main forces of protest for fear of inciting a full-scale revolution. At the same time, as many demonstrators announced in public on various occasions, the protest was against the regime, not for the opposition party.

When we examine protest in the context of national movements, however, we may find that the general political circumstances and the sense of solidarity in opposition to the regime, nurtured by the national desire for change, made an important contribution to the origins and dynamics of protest movements. Just as subscribing to a newspaper with a certain orientation binds the readers into a symbolic community of political opinion (Alexander 1988), a history of similar voting patterns in different places at the national level makes it easier for potential organizers and participants to initiate protest because they believe that their call for change will be answered by persons in other cities.

One of the most important institutional features of the time that helps us understand the findings is the Rhee regime's authoritarian and repressive nature. An authoritarian regime cannot survive massive protest movements without intensifying its repression and a repressive regime often collapses at a point when the effectiveness of repression reaches a certain threshold of not being able to exercise its coercion on the citizens. A democratic state is stronger than an authoritarian regime because it has a broad basis of support from the people. An authoritarian regime usually has a very narrow basis of support and when the opposition to it becomes widespread, it is doomed to collapse. As the size of the opposition grows, the legitimacy of the opposition also gains root. In the post-war period in South Korea, an irreversible political trend developed under the structural influence of political authoritarianism and economic stagnation. The more the repression became intensified, the more widespread the opposition to the regime grew in the years prior to the upheaval.

In the case of the 4-19 Revolt, a concatenation of special circumstances produced a critical situation: a high concentration of the unemployed and students and of potential participants and supporters in major cities, especially Seoul, the capital city; unity among the major elements of the political opposition; and critical public opinion.

Because of the catastrophe caused by the Korean War, the government was fiscally bankrupt and poorly prepared for the massive upheaval. It became morally bankrupt and could not defend its position under the mantle of law and order when it succeeded in stealing an election while drawing its legitimacy from democratic, constitutional principles. The U.S. government promoted democracy abroad, especially among dependent states such as South Korea, but was embarrassed by the blatant disregard for democracy on the part of Rhee and his party leaders. The Eisenhower administration condemned the rigging and blamed the leaders of South Korea for their failure to take responsibility for

the misrule. The timely announcement of condemnation encouraged the opposition forces to demand President Rhee's resignation. Therefore, when students, intellectuals, and university professors — including some famous nationalists — took the lead in protesting, the government could not define the dissidents as criminals or frame them as communists and thus crush the movement by legal-rational authority.

The leadership of the movement by students in elite universities legitimated the revolt, and the mass media championed the cause of protest. The authorities found it difficult to control the revolt because the protesting groups were diverse, disconnected, complex, and large, without a unified command. If a single national leadership had existed, the authorities could have destroyed it easily. As if shooting in the dark, however, the elites responded to the protest in an indiscriminate, incoherent, and unusually harsh manner. This combination of harshness of policy and weakness of legitimacy allowed the protest movement to become an open rebellion against the regime. In fact, ironic as it may seem, the government could not control the rebellion by force because it used force and thus undermined its own legitimacy.

Violent coercion, especially reactive rather than proactive coercion, transforms protest into revolutionary upheaval. Violent coercion, when exercised extensively and preemptively, may deter protest and the expansion of protest. Insufficiently intense coercion, however, when exercised in reaction to protest may be ineffective. Violent coercion becomes counterproductive when the elites are divided among themselves about the use of such coercion and when the protest against which it is intended gains popular support. Furthermore, coercion exposes the contradictions between the principles and the practices of the regime in an authoritarian state. A movement becomes revolutionary by breaking down the coercive power, or by making it difficult for the elite to use force in controlling the movement, because the elite has destroyed the moral legitimacy of the use of force.

By forcing the leaders out of power, the 4-19 Revolt destroyed the rules and procedures of the authoritarian regime. In doing so, it revived the principles and norms of the state from which the regime drew its legitimacy, and opened the way for democracy.

This study offers a theoretical model that suggests that protest movements develop into revolutionary movements that forces a change of regime when protest occurs under conditions of extensive urban unemployment combined with a history of opposition to the regime. The regime will collapse under such a pressure when it uses coercion against the movement that is based on a widespread critical public opinion. Insofar as these conditions interact in such a way as to politicize economic deprivation, to increase participation of citizens across different cities, and to increase legitimacy of the opposition while undermining the legitimacy and effectiveness of the use of coercive force, then the repressive and authoritarian regime will collapse. Such a model may not fit political processes of change of regime across different cultures and historical periods. Different structural conditions produce different varieties of change of regime. However, the historical conjunctures and specific circumstances associated with change of regime discovered here developed from political conditions and trends similar to those one can commonly observe in many of

the developing societies in the contemporary world. As a research question, why some regimes change under similar conditions while others do not presents a challenge to all students of protest and revolutionary movements.

Notes

1. The event that we study can be defined broadly not as revolution but as revolt. In revolution, the transfer of power as an outcome of a revolutionary situation is more extensive than in revolt. Protest becomes a revolutionary movement, a revolutionary upheaval or a rebellion as the demand for change shifts from issues related to policies and procedures to issues concerning principles, norms, and leadership of the regime (Tilly 1978, 1993). Hibbs (1973), Gurr (1980), and Gurr and Lichbach (1986) distinguish protest from rebellion for analytical purposes and find in cross-national comparisons that these events have different causes and consequences.

The distinction between protest and revolutionary upheaval or rebellion that I use here is similar to that proposed by Jenkins and Schock (1992): "Social protests [are] aimed at limited issues such as changing the policies of authorities or particular personnel; and rebellions [deal] with fundamental issues such as who governs and what is the structure of authority" (162). For more extensive definitions of protest, see Eckstein (1989: xi) and Lofland (1985:1-2). For broader definitions of revolution, see Skocpol (1979:4). I use the terms, *revolutionary movements, revolutionary upheaval,* and *rebellion* interchangeably to describe collective actions in revolutionary situations.

2. Deprivation as a social condition has both synchronic and diachronic implications; thus deprivation, by definition, is relative to time (Davies 1962) and to a comparative frame of reference (Gurr 1970). Though deprivation is common and widespread throughout much of human experience, it is not constant but varies according to time and place; thus relevant to an explanation of collective protest.

3. Results from cross-national studies, however, are generally weak for this contention and are inconclusive at best (see Gurr & Duvall 1973; Hibbs 1973; Zimmermann 1980).

4. Reported by *Dong-A Ilbo* (3-20 March, 1960) and documented later by *The History of the Korean Revolutionary Trials* (Korean Government 1962).

5. According to a subsequent report, seven persons were killed and 72 were injured (see J. Oh 1968:61).

6. Interview with Chong-Sik Chong on 22 June, 1960 (in D. Kim & Ahn 1960).

7. Syngman Rhee left Korea in May for Hawaii, where he died in exile in fall 1963. An interim government under Ho Chong conducted a new election for the National Assembly, from which Chang Myon emerged as the leader. Unable to institutionalize political power (Han 1974), Chang Myon succumbed in May 1961 to a military *coup d'etat* led by General Park Chong-Hee.

8. This strategy is similar to what Snyder (1976) called "event linkage," whereby one attempts to predict the characteristics of one event from the characteristics of the previous events.

9. The strength of opposition to the regime also can be gauged by a proportional measure — for instance, by dividing the percentage of the opposition vote by the percentage of the vote for the ruling party. This measure, however, overestimates the extent of dominance by the opposition and underestimates the support for the ruling party. For example, the index will be .111 in a city where the vote is 10% for the opposition and 90% for the ruling party, 1.00 when the vote is divided evenly, and 9.00 when the pattern is reversed, with 90% for the opposition and 10% for the ruling party. Because I am more interested in measuring the pattern of power difference than of power dominance, I subtract the vote for the ruling party from the opposition vote. This measure reflects the extent of power difference equally for the opposition and the ruling party. In any case, the measure of difference is correlated highly with the measure of dominance ($r = .997$) in this study.

APPENDIX A: Cities and Measures Used for Analysis[a]

City	(1) Protest 1 (Log)	(2) Violent Coercion	(3) Protest 2 (Log)	(4) Unemploy-ment Rate (%)	(5) Newspaper Reader-ship (%)	(6) Opposi-tion Vote (%)	(7) Student Population (%)
Seoul	6.49	85.89	10.16	18.7	45.7	95.5	29.4
Pusan	5.26	15.24	6.17	21.2	44.6	58.1	19.5
Taegu	4.88	1.28	5.38	13.1	40.2	84.6	17.3
Inchon	3.56	0.85	5.12	18.7	27.3	68.5	15.2
Kwangju	1.39	17.05	3.69	11.6	34.0	-2.7	20.7
Taejon	3.74	1.04	4.97	15.2	30.3	46.7	18.1
Chonju	1.79	1.08	4.14	16.0	33.2	82.4	18.8
Masan	8.83	90.02	6.70	26.2	29.1	85.8	19.2
Mokpo	1.79	0.90	2.71	15.1	22.8	75.0	20.5
Chongju	5.05	1.49	3.58	18.0	26.1	33.9	21.9
Suwon	3.47	0.90	3.18	20.0	25.0	66.2	17.5
Kunsan	1.39	0.60	4.84	19.4	30.6	25.8	17.0
Yosu	2.30	0.70	0.69	19.5	12.0	-6.2	14.6
Chinju	3.69	0.85	2.77	12.0	15.6	15.2	18.5
Chunchon	1.39	0.30	2.10	15.4	29.5	20.2	15.9
Kyongju	1.39	0.30	3.00	9.1	21.7	45.0	14.1
Sunchon	1.39	0.30	3.00	7.8	16.9	36.4	16.7
Chungju	2.71	0.70	2.49	9.4	22.1	-8.9	14.4
Chinhae	2.08	0.30	0.69	12.3	21.0	61.0	14.0
Cheju	1.79	0.30	2.08	4.7	20.8	-12.2	18.3
Iri	1.79	0.48	4.99	15.2	36.8	22.9	19.6
Pohang	3.00	1.00	6.17	15.1	16.2	-39.0	5.3
Kimchon	1.79	0.48	4.09	10.8	30.6	63.4	17.4
Mean	3.08	9.65	4.00	15.0	27.5	42.4	8.0
Std. dev.	1.93	25.11	2.08	5.0	9.0	37.1	3.3

[a] Sources: (1), (2), and (3): see text; (4) Bureau of Statistics, Ministry of Home Affairs (1961), Economic Planning Board (1961); (5) Ministry of Public Information (1961); (6) Central Election Management Commission (1971); (7) Education Press (1961).

APPENDIX B: Correlation Matrix of Variables Used for Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Unemployment rate	—						
(2) Newspaper readership	.335	—					
(3) Opposition vote	.352	.443*	—				
(4) Student population	.269	.599**	.381	—			
(5) Protest 1	.632**	.374	.466*	.466*	—		
(6) Violent coercion	.500*	.415*	.388	.625**	.762**	—	
(7) Protest 2	.468*	.723**	.398	.670**	.651**	.693**	—

* p < .05 ** p < .01 (two-tailed test)

References

Alexander, Jeffrey C. 1988. *Action and Its Environments*. Columbia University Press.

Aya, Rod. 1984. "Popular Intervention in Revolutionary Situations." Pp. 318-43 in *Statemaking and Social Movements*, edited Charles Bright and Susan Harding. University of Michigan Press.

Bureau of Statistics, Ministry of Home Affairs. 1961. *Monthly Bulletin of Statistics*, no. 1. Seoul: Bureau of Statistics.

Central Education Research Institute. 1967. *A Comprehensive Study of Higher Education*. Seoul: Central Education Research Institute.

Central Election Management Commission. 1971. *Presidential Election Patterns*. Seoul: Central Election Management Commission.

Cho, Jong-Kuk. 1969. "A Content Analysis of *Chosun Ilbo* Editorials." Master's thesis, Institute of Mass Communication, Seoul National University.

Coleman, James S. 1990. *Foundations of Social Theory*. Harvard University Press.

Davies, James C. 1962. "Toward a Theory of Revolution." *American Sociological Review* 27:5-19.

DeNardo, James. 1985. *Power in Numbers: The Political Strategy of Protest and Rebellion*. Princeton University Press.

Eckstein, Susan. 1989. "Power and Popular Protest in Latin America." In *Power and Popular Protest*, edited by Susan Eckstein. University of California Press.

Economic Planning Board. 1961. *1960 Population and Housing Census of Korea*. Vol. 1. Seoul: Economic Planning Board.

Education Press. 1961. *Korea Education Almanac*. Seoul: Education Press.

Granovetter, Mark. 1978. "Threshold Models of Collective Behavior." *American Journal of Sociology* 83:1420-43.

Gupta, D.K. 1977."Socio-Economic Costs of Unemployment and Income Inequality: A Cross-National Study, 1948-1967." Ph.D. diss., University of Pittsburgh.

Gurr, Ted Robert. 1970. *Why Men Rebel*. Princeton University Press.

——. 1980. "On the Outcomes of Violent Conflict." Pp. 238-94 in *Handbook of Political Conflict*, edited by Ted Robert Gurr. Free Press.

Gurr, Ted Robert, and R. Duvall. 1973. "Civil Conflict in the 1960's: A Reciprocal Theoretical System with Parameter Estimates." *Comparative Political Studies* 6:135-70.

Gurr, Ted Robert, and Jack A. Goldstone. 1991. "Comparisons and Policy Implications." In *Revolutions of the Late Twentieth Century*, edited by Ted Robert Gurr, Jack A. Goldstone, and Farrokh Moshiri. Westview.

Gurr, Ted Robert and Mark Irving Lichbach. 1986. "Forecasting Internal Conflict: A Competitive Evaluation of Empirical Theories." *Comparative Political Studies* 19: 3-38.

Han, Sung-Joo. 1974. *The Failure of Democracy in South Korea*. University of California Press.

Hibbs, Douglas A. Jr., 1973. *Mass Political Violence: A Cross-National Causal Analysis*. Wiley.

Horwitz, Allan V. 1984. "The Economy and Social Pathology." *Annual Review of Sociology* 10:95-119.

In, Bo-gil. 1970. "Content Analysis of *Dong-A Ilbo*, 1945-60." Master's thesis, Institute of Mass Communication, Seoul National University.

Jackman, Robert W. 1993. *Power without Force: The Political Capacity of Nation-States*. University of Michigan Press.

Jacobs, Norman. 1985. *The Korean Road to Modernization and Development*. University of Illinois Press.

Jenkins, J. Craig. 1983. "Resource Mobilization Theory and the Study of Social Movements." *Annual Review of Sociology* 9:527-53.

Jenkins, J. Craig, and Kurt Schock. 1992. "Global Structures and Political Processes in the Study of Domestic Political Conflict." *Annual Review of Sociology* 18:161-85.

Kim, Dalchoong, and Byung-Joon Ahn. 1960. *The Research Material of the April Revolution*. Seoul: Yonsei University Library.

Kim, Jongchol. 1985. *Education and Development*. Seoul: Seoul National University Press.

Kim, Kyung-Dong. 1959. "Sociological Studies of the Hope and Value of the Rural Youth." *Seoul National University Nonmunjip* 7:30-42.

Kim, Quee-Young. 1983. *The Fall of Syngman Rhee*. University of California, Berkeley: Institute of East Asian Studies.

____. 1991. "Disjunctive Justice and Revolutionary Movements." Pp. 56-70 in *Revolutions in the Third World*, edited by Quee-Young Kim. Leiden: Brill.

Kim, Song-Tae. 1960. "A Study of the Motivating Factors of the April 19th Student Upheaval." *Songkyunkwan Nonmunjip* 5:80-106.

Kitschelt, Herbert P. 1986. "Political Opportunity Structures and Political Protest: Anti-Nuclear Movements in Four Democracies." *British Journal of Political Science* 16:57-85.

Korean Government. 1962. *The History of the Korean Revolutionary Trials*. Seoul: Korean Government.

Krasner, Stephen D. (ed.). 1983. *International Regimes*. Cornell University Press.

Lawson, Stephanie. 1993. "Conceptual Issues in the Comparative Study of Regime Change and Democratization." *Comparative Politics* 25:183-206.

Lee, Ki-Baik. 1984. *A New History of Korea*, trans. E. Wagner. Harvard University Press.

Lofland, John. 1985. *Protest: Studies of Collective Behavior and Social Movements*. Transaction Books.

Masan Daily. 1960. "Analysis of the Masan Upheaval." 3 June, pp. 3-4.

Ministry of Public Information, Government of Korea. 1961. *Newspaper Distribution Patterns in Korea* (in Korean). Seoul: Ministry of Public Information.

Molotch, Harvey. 1979. "Media and Movements." Pp. 71-93 in *The Dynamics of Social Movements*, edited by Mayer N. Zald and John D. McCarthy. Winthrop.

Moore, Barrington, Jr. 1978. *Injustice: The Social Bases of Obedience and Revolt*. Sharpe.

Muller, Edward N. 1972. "A Test of a Partial Theory of Potential for Political Violence." *American Political Science Review* 66:928-59.

____. 1979. *Aggressive Political Participation*. Princeton University Press.

____. 1986. "Income Inequality and Political Violence: The Effect of Influential Cases." *American Sociological Review* 51:441-45.

Muller, Edward N., Henry A. Dietz, and Steven E. Finkel. 1991. "Discontent and the Expected Utility of Rebellion: The Case of Peru." *American Political Science Review* 85:1261-82.

Oberschall, Anthony. 1973. *Social Conflict and Social Movements*. Prentice-Hall.

____. 1978. "Theories of Social Conflict." *Annual Review of Sociology* 4:291-315.

____. 1980. "Loosely Structured Collective Conflict." Pp. 45-68 in *Research in Social Movements, Conflict and Change*, edited by Louis Kriesberg. JAI.

Oh, Byung-Hun. 1975. "Students and Politics." Pp. 111-52 in *Korean Politics in Transition*, edited by Edward Reynolds Wright. University of Washington Press.

Oh, John Kie-Chiang. 1968. *Korea: Democracy on Trial*. Cornell University Press.

Oliver, Pamela, and Gerald Marwell. 1988. "The Paradox of Group Size in Collective Action: A Theory of Critical Mass. II." *American Sociological Review* 53:1-8.

Oliver, Pamela, Gerald Marwell, and Ruy Teixeira. 1985. "A Theory of Critical Mass. I: Interdependence, Group Heterogeneity, and the Production of Collective Action." *American Journal of Sociology* 91:522-56.

Oliver, Robert T. 1957. "Present-Day Newspapers in the Republic of Korea." *Journalism Quarterly* (Winter):85-87.

Pak, Chi-Young. 1980. *Political Opposition in Korea, 1945-1960*. Seoul: Seoul National University Press.

Pareto, Vilfredo. [1901] 1991. *The Rise and Fall of Elites*. Transaction Books.

Sa-Il-Gu Yujokhoe. 1971. *A Commemorative Publication for the Tenth Anniversary of the Sa-Il-Gu*. Seoul: *Sa-Il-Gu* Yujokhoe.

Scalapino, Robert A. 1961. "Korea." In *United States Foreign Policy — Asia*, prepared by Colon Associates, Ltd. under the direction of the Committee on Foreign Relations, 87th U.S. Congress. Government Printing Office.

Skocpol, Theda. 1979. *States and Social Revolutions*. Cambridge University Press.

Snyder, David. 1976. "Theoretical and Methodological Problems in the Analysis of Governmental Coercion and Collective Violence." *Journal of Political and Military Sociology*. 4:277-93.

Snyder, David, and Charles Tilly. 1972. "Hardship and Collective Violence in France, 1830-1960." *American Sociological Review* 37:520-32.

Stinchcombe, Arthur L. 1965. "Social Structure and Organizations." In *Handbook of Sociology*, edited by James G. March. Rand McNally.

Tilly, Charles. 1974. "The Chaos of the Living City." Pp. 86-108 in *An Urban World*, edited by Charles Tilly. Little, Brown.

____. 1978. *From Mobilization to Revolution*. Addison-Wesley.

____. 1979. "Repertoires of Contention in America and Britain, 1750-1830." In *The Dynamics of Social Movements*, edited Mayer N. Zald and John D. McCarthy. Winthrop.

____. 1986. *The Contentious French*. Belknap.

____. 1989. "Collective Violence in European Perspective." Pp. 62-100 in *Violence in America*, edited by Ted Gurr. Sage.

____. 1993. *European Revolutions, 1492-1992*. Blackwell.

Walton, John. 1984. *Reluctant Rebels*. Columbia University Press.

Walton, John, and Charles Ragin. 1990. "Global and National Sources of Political Protest: Third World Response to the Debt Crisis." *American Sociological Review* 55:876-90.

Zimmermann, Ekkart. 1980. "Macro-Comparative Research on Political Protest." In *Handbook of Political Conflict*, edited by Ted Robert Gurr. Free Press.

Nonconformity in Competitive Repertoires: A Sociological View of Markets*

DANNY MILLER, *École des Hautes Études Commerciales, Montreal and Columbia University*
MING-JER CHEN, *Columbia University*

Abstract

A sociological framework is proposed for studying nonconformity in competitive repertoires. We argue that the extensiveness and continuity of a company's market interactions very much influence its level of nonconformity. Market-level factors that might influence nonconformity include customer and competitor diversity and the continuity of traditions; and firm-level factors include the size, resources, and industry experience of an organization. Nonconformity may also be driven by economic factors such as market growth and financial returns. Most of these sources of nonconformity were confirmed in a study of the competitive repertoires of the major U.S. airlines. The findings also show that competitive nonconformity was associated with subsequent declines in financial performance, but mainly for smaller airlines with homogeneous markets. This study supported some central notions from the sociology of markets and, with qualification, from institutional theory.

Some organizations are conformists. They compete in conventional ways and act much like their competitors. Others are iconoclasts and adopt competitive practices that deviate from industry norms. This study looks into some of the factors that underlie this nonconformity. Although we examine economic sources of nonconformity, we will pay particular attention to those influences suggested by the sociology of organizations, markets, and networks, viewpoints that have been largely ignored by scholars of competitive strategy.

The issue of nonconformity could well prove important. Companies deviating from conventional modes of competition may surprise their rivals (Chen & MacMillan 1992), develop appeals that are distinctive and enduring (Porter 1985; Prahalad & Hamel 1990), and be difficult to imitate and retaliate

** We would like to thank Richard Harrison, Heather Haveman, Kuo-Hsien Su, and two fine reviewers for their useful comments on an earlier version of the article. The first author is also indebted to the Social Sciences and Humanities Research Council of Canada and to École des Hautes Etudes Commerciales for partial funding of this research. The second author acknowledges Columbia Business School for its partial funding of this research. Direct correspondence to Danny Miller, 4642 Melrose Ave., Montreal, P.Q., Canada, H4A 2S9.*

against (Chen & Miller 1994). But nonconformists may fail to supply customers with the products or services that they have come to expect. Indeed, some institutional sociologists have argued that nonconformity can erode organizational legitimacy as perceived by clients and regulators (DiMaggio & Powell 1991; Scott 1987). Also, unconventional rivalry may risk evoking hostile reactions from competitors (Chen & Miller 1994).

Scope of the Study

This study compares "conventional" or isomorphic competitive repertoires to "nonconformist" ones that depart more markedly from industry central tendencies or de facto norms. Competitive repertoires consist of the entire set of market-oriented actions used by individual firms to attract customers and cope with rivals (Miller & Chen 1994). These actions may include price changes, new product and service introductions, and geographic expansions. Conventional repertoires employ a mix of actions typical to firms of similar scope and scale in the same industry. Nonconformist repertoires, on the other hand, encompass a set of actions that are atypical — that emphasize market-oriented ploys rarely used by competitors or that avoid actions frequently used by rivals.

Conventional or isomorphic behavior has been of great interest to institutional sociologists, some of whom view conformity as an important way for organizations to gain legitimacy and cooperation from regulators, competitors, and even customers (Powell 1991; Scott 1987). Conformity has also been of central concern to sociologists studying the emergence and diffusion of practices and norms in groups, networks, and markets (Burt 1987; Granovetter 1985). In using the sociological literature to derive our framework, we take as our point of departure White's (1981) dictum that markets are "social structures among specific cliques of firms and other actors who evolve roles from observations of each other's behavior. [What] a firm does in a market is to watch the competition in terms of observables" (518). Indeed, the thesis of this study is that market conformity and nonconformity in the competitive repertoires of organizations can best be analyzed from a sociological perspective that departs fundamentally from economic approaches to business strategy. More specifically, we will argue that *nonconformity will be very much a product of an organization's social context and its interactions with and relationship to that context.*

The focus of this study, however, will differ from most of the earlier sociological literature. First, as indicated above, we examine the entire repertoire of a company's market-oriented competitive actions, not just the unitary rituals or administrative adoptions that are the typical foci of most studies (Burt 1987; Zucker 1987). Second, we do not posit any a priori norms of tactical conduct, since they are absent, unclear, or ephemeral in most competitive markets (Levine 1987). Instead, our concern is with de facto norms — that is, with patterns of competitive behavior that are typical of a group of interacting competitors of similar market scope. Nonconformity can be assessed relative to such typical behavior. Third, we do not view conformity or nonconformity as an imperative, equilibrium condition, or binary choice, but rather as a continuum along which competitors range widely. Fourth, our model incorporates

economic as well as sociological sources of nonconformity. This is done to arrive at a more complete predictive model and to show the complementarities among approaches. Finally, we assess the financial implications of competitive nonconformity at the organizational level. These traditionally have been of more concern to economists and strategy theorists than to sociologists (Mezias 1990; Oliver 1988).

Sources of Competitive Nonconformity

A FRAMEWORK FOR STUDYING COMPETITIVE NONCONFORMITY

It has been suggested that organizations often gravitate toward administrative isomorphism (Fligstein 1985; Haveman 1993b; Palmer, Jennings & Zhou 1993; Spender 1989). Convergence upon a "conventional" competitive repertoire may be common as well. When geographic expansion becomes popular in an industry, some executives feel obliged to get into the act (Fligstein 1985; Spender 1989). Where price cutting is common, many managers are induced to respond in like fashion (Chen 1988). Scholars of business strategy have attributed mostly rational economic motives to such behavior (Porter 1980, 1985; Rumelt, Schendel & Teece 1991). And although we do not rule out such motives, we believe that they are very much conditioned by an actor's social relationships (Granovetter 1985; White 1981, 1988). More specifically, the extensiveness and continuity of interaction of an organization with other actors in its market will help explain its conformity to de facto industry norms.

The social context of an organization can have a great impact on conformity. Some contexts, for example, encourage *extensive* interaction: that is, diverse relationships with a wide variety of actors. Such interactions expose members to diverse perspectives, encourage independent thought, and are quite conducive to nonconformity (Granovetter 1973, 1985). Other contexts, by contrast, are characterized by contacts with a rather homogeneous group of actors. The social cohesion of such homogeneous environments typically exerts strong pressures for conformity (Burt 1987; Homans 1950).

Another important aspect of the context that bears on nonconformity is the *continuity* or duration of traditions. Some contexts are in a state of upheaval and have not yet settled down enough to allow the emergence of norms. Others are more established and imbued with tradition. Clearly, nonconformity will be strongest where norms have not yet been firmly established (Erickson 1988; Homans 1950).

Nonconformity, of course, is not simply a function of an organization's social context but also its links to, dependencies on, and embeddedness within that context. And again the extensiveness and continuity of interaction play important roles. Large firms typically have extensive dealings with a great many kinds of rivals and customers, some of whom do things rather unconventionally, and none of whom the giants are much dependent upon. Many large firms thus have the broad experience, the freedom, and the resources to deviate from industry norms (Haveman 1993a). Smaller firms, on the other hand, are more likely to find themselves interacting with a few similar actors upon whom they

depend a great deal. Convention is hard to ignore under these conditions (DiMaggio & Powell 1983, 1991; Granovetter 1985).

If the continuity of industry traditions encourages conformity at the market level, then the continuity of industry experience plays a socializing role at the firm level. Older, experienced companies are likely to have been more thoroughly exposed to and influenced by industry traditions than are new entrants to an industry (Meyer & Zucker 1989). This will dampen their nonconformity.

As important as these social factors are, we expect them to be complemented by two key economic considerations. Specifically, the lure of opportunities and the threat of poor returns may cause companies to move away from conventional modes of competing (Porter 1980:215-35; Scherer 1980:407-58; Schumpeter 1950). At the market level, some of these influences may take the form of growth possibilities, and at the firm level, they may appear as lapses in financial performance.

The sources of nonconformity are arrayed in Table 1.

THE SOCIAL CONTEXT

The impact of group norms as well as an actor's ability to entertain possibilities that extend beyond those norms will be influenced by the continuity and extensiveness of social interaction. These conditions are, in part, a function of the market context within which a firm operates: industry traditions and the diversity of markets being especially important.

The Diversity of Markets

The perceived realities of administrators are very much influenced by their interactions with their environments. The originality of managerial viewpoints will depend in part on whether firms have most of their relationships with a small group of actors — that is, customers and competitors — or whether they spread these relationships more broadly (Berger & Luckmann 1967; Scott 1987). Burt (1987) and Erickson (1988) suggest that actors are more likely to conform if they depend on a tightly networked clique of similar others. Conversely, "the socializing influence of cohesive ties within a clique can be eliminated by conflicting ties outside the clique" (Burt 1987:1293). Indeed, Granovetter (1973) shows there to be an inverse relationship between the intensity and extensiveness of relationships: actors with very intensive ties to a few parties tend not to have very extensive ties, and vice versa. In any event, extensive and diverse interactions may dampen the salience of norms and encourage nonconformity.

Certainly, these observations may apply to two central aspects of a market environment: diversity in the set of competitors and diversity in the set of customers. Firms that compete against many different rivals for numerous market niches are apt to receive more unconventional ideas for competing than firms that interact with a more homogeneous set of contacts (Burt 1987; Erickson 1988). The former are exposed to a wider variety of competitive challenges that inform them of unorthodox ways of conducting business (Miller 1993; Miller & Chen 1993). A diverse set of customers is also likely to generate demands for unusual services and hence for nonconformity.

TABLE 1: Sources of Nonconformity

	Extensiveness of Interaction	Continuity of Interaction	Economic Motivations
Social context (market level)	Customer and competitor diversity	Absence of industry traditions	Market growth
Links with context (firm level)	Firm size and resources	Firm inexperience	Poor returns

Companies that compete for the same customers and niches of the environment are dependent for resources and acceptance on a relatively small group of actors (rivals, clients, institutions), and so are quite susceptible to pressures from them to conform (Homans 1950; Pfeffer & Salancik 1978; Powell 1991). By contrast, organizations with more extensive interactions and a vaster market scope may be sufficiently independent to resist such pressures (Scott & Meyer 1987).

Hypothesis 1. Competitive nonconformity will be positively related to the diversity of a firm's competitors and customers.

Industry Traditions: The Time since Market Upheaval

Erickson (1988) and Homans (1950) showed that conformity increased with the continuity and stability of contact among group members. By contrast, an absence of continuity or experience in interaction, as prevails in very young industries or during periods of upheaval, precludes convention and tradition, and thus provides the latitude for nonconformity (DiMaggio 1991; DiMaggio & Powell 1991; Granovetter 1985).

In new industries or in those undergoing a thorough upheaval, there are no established rules of the game (Boeker 1989; Powell 1985; Tushman & Anderson 1988). A heterogeneous mix of competitive strategies are conducted as firms collectively grope their way toward a viable approach. Here there are few norms or traditions to serve as guides and few customer expectations or standards with which to comply (Mezias 1990). But in time, implicit norms and conventions develop. Firms begin to conform to certain "accepted" ways of competing (DiMaggio 1991; DiMaggio & Powell 1983; Spender 1989). The passage of time, in other words, brings with it mutual mimesis and the mantle of tradition (DiMaggio 1991; Fligstein 1985; Palmer, Jennings & Zhou 1993; Powell 1985).

Hypothesis 2. Competitive nonconformity will be negatively related to the time since the last major industry upheaval.

AN ORGANIZATION'S LINKS TO THE SOCIAL CONTEXT

Although the social context can have a profound impact on competitive nonconformity, a given firm's dependence upon and embeddedness within that context may be important as well. The size and resources of a company, for example, may give it the extensive experience and the means to enjoy considerable independence from the pack. Conversely, the age of an organization may reflect the depth of its exposure to the socializing forces of industry tradition.

Firm Size and Slack Resources

The role of company size and resources in stimulating originality seems to pull in opposite directions. On the one hand, large and rich organizations typically are exposed to many different ways of competing (Burt 1987). Also, their scale makes them less dependent and thus less susceptible to the conforming influences of a few key relationships (Granovetter 1973). Many such large enterprises can afford to take risks and stray from convention (Haveman 1993a). And their slack resources may fund the development of more elaborate and original repertoires (Cohen & Levinthal 1990). Moreover, large and wealthy firms may be able to buffer themselves from external threats (Pfeffer & Salancik 1978) and avoid complying with industry norms (Mezias 1990).

On the other hand, many institutional and bureaucratic theorists would argue that size and resources have a dampening effect on nonconformity. Some large organizations adhere to convention simply because they are so visible to — or so closely monitored by — regulators (Cooper, Willard & Woo 1986; Jacobs, Useem & Zald 1991; Powell 1991; Zelditch & Walker 1984; Zucker 1987). Size may also encourage bureaucracy and inertia (Hannan & Freeman 1984; Haveman 1993a). Smallness, by contrast, forces some companies to be original, since they have not the resources to vie with their larger rivals using conventional approaches.

Conflicts between network and resource dependency views on the one hand, and institutional and bureaucratic arguments on the other, highlight the tentativeness of the following hypothesis. However, in industries like the one we have studied, big and small firms face similar regulatory scrutiny, and considerable resources are needed for many kinds of originality. Under these circumstances we tend to favor the network and resource dependency arguments.

Hypothesis 3. Competitive nonconformity will be positively related to the size of the firm and its level of slack resources.

Firm Age

If, as we have argued, industry traditions can generate conformity, then so can a firm's continuity of exposure to those traditions. This, in part, is determined by the age of an organization (Aldrich & Auster 1986; Miller 1993; Tushman & Romanelli 1985). Organizations that have long been exposed to typical industry practices and norms are that much more likely to be influenced by them. Meyer & Zucker (1989) argued that well-established organizations develop such stable ties to other institutions, businesses, and regulatory bodies that their ability to stray from contemporary norms is limited. These alliances buffer and protect

firms from external challenges, thereby making nonconformity still less likely (Burt 1987; Granovetter 1973). Young firms, by contrast, are more apt to be nonconformists, in part because of their lack of experience with industry traditions and in part because of their rather modest institutional attachments (Baum & Oliver 1992; Stinchcombe 1965).

As corporations get older, they also become wed to traditional ways of doing things, to more automatic forms of behavior, and to learning that refines conventional approaches instead of developing original ones (March 1991; Starbuck 1983). Anything unorthodox is viewed with suspicion as being too risky or as negating cherished traditions. Deviations are also seen to challenge the past contributions of those in power (Staw 1981).

Hypothesis 4. Competitive nonconformity will be negatively related to the age of the firm.

ECONOMIC INCENTIVES FOR NONCONFORMITY

So far we have discussed purely sociological determinants of nonconformity. But there are at least two major economic incentives for nonconformity: the pursuit of growth opportunities and the avoidance of poor returns. As we shall see, however, even these economic considerations are best understood within their social contexts.

Market Growth

Managers are apt to depart from traditional ways of doing things during periods of environmental expansion — when the growth of an industry or the emergence of new market niches encourage unorthodox means of capturing additional business. During periods of rapid growth, companies experiment with a variety of techniques to appeal to the burgeoning army of potential customers (Chandler 1962; Scherer 1980; Tushman & Anderson 1988). Moreover, the new rivals and customers that enter markets during periods of growth may erode competitive conventions and elicit unorthodox new methods (Porter 1980:215-33; Schumpeter 1950). Rapid market growth also fosters managerial confidence that can again motivate departures from convention. The blossoming years of the semiconductor industry provides a good example of such competitive creativity (Eisenhardt 1989).

Eras of slow growth or stagnation, by contrast, often encourage competitive isomorphism. An unchanging stock of customers and rivals allows conventional wisdom to dominate (DiMaggio 1991; Zucker 1987). Moreover, in stagnating contexts there is little incentive to chase after customers using unconventional approaches as there is not much new business to be had. Stagnant contexts are also pervaded by a shortage of resources, which renders firms more susceptible to the economic and institutional pressures imposed by powerful stakeholders (DiMaggio & Powell 1983, 1991; Pfeffer & Salancik 1978:39-52). And because corporate conservatism often accompanies a downturn, expensive forms of nonconformity are discouraged.[1]

Hypothesis 5. Competitive nonconformity will be positively related to market growth.

A Firm's Poor Performance

In light of the above it is perhaps paradoxical that threat can be as great a stimulus for nonconformity as market opportunity. Yet scholars of organizations note that companies — the majority of whom, by definition, follow conventional repertoires — are likely to abandon conventional repertoires mainly under pressure of necessity. Abysmal returns to capital are great incentives to search for novel remedies (Amburgey & Miner 1992; Cyert & March 1963). Poor performance, at least in the majority of firms that are conforming, motivates managers to try something less orthodox (Lant & Montgomery 1987; Starbuck, Greve & Hedberg 1978). Even the literature on prospect theory argues that the propensity to take risks — and therefore, perhaps, to be deviant — increases in the domain of losses (Tversky & Kahneman 1973).

Good performance, on the other hand, fosters complacency and encourages managers to take the path of least resistance (Miller 1993; Milliken & Lant 1991). According to institutional and bureaucratic scholars, this attitude, more often than not, means sticking with standard, "tried and true" industry recipes (Lant & Montgomery 1987; Spender 1989). Nonconformist strategies are often risky, they are difficult to devise, and they may induce political resistance from inside and outside an organization (Scott 1987). Hence they are unlikely to be used by successful firms whose managers are not being pressured to undertake such tribulations. Heterodoxy, in short, typically demands greater incentives than conformity.[2]

Hypothesis 6. Competitive nonconformity will be negatively related to prior organizational performance.

The Hazards of Nonconformity

Competitive nonconformity, we expect, will usually hurt financial performance because it increases costs and competitive risks and reduces the acceptability of organizational actions. Moreover, nonconformity will be especially inappropriate to small firms operating in undiversified markets as risks of failure there are high. Large organizations, on the other hand, may have the resources to make deviant repertoires acceptable to their customers; and diversified firms may be able to reduce the risks of nonconformity and use it to good advantage in those of their markets that depart from industry norms.

The strategy/industrial economics literature and the writings on institutional sociology seem to be at odds over the implications of competitive nonconformity. Porter (1980, 1985) has espoused the merits of distinctive (hence nonconformist) competitive repertoires that hit rivals where they are weakest. For example, he suggests differentiating products by constant quality or image improvements or developing unique cost advantages that allow frequent price-cutting. And indeed some such unconventional tactics have the advantage of being difficult to imitate (Lieberman & Montgomery 1988). Others are hard for competitors to detect and respond to because they are stealthy (Chen & MacMillan 1992; Chen & Miller 1994). And still others enable firms to develop

distinctive resources and competitive advantages (Porter 1985:120-64; Prahalad & Hamel 1990). But nonconformity in and of itself is unlikely to generate economic rents unless it meets the needs of customers and the challenges of competitors. And therein lies its major difficulty. Unique products and processes may do more to scare off customers and stir up rivals than to reduce the firm's vulnerability.

Thus the dangers of nonconformity may well outweigh its advantages. The defiance of convention often carries with it hazards of novelty and unpredictability, the possibility of attracting attention from aggressive imitators, and the elevated costs of convincing customers of the advantages and legitimacy of an unusual approach (Kreps 1990; Shamsie 1989; Wernerfelt & Karnani 1987).

Institutional theorists, in particular, have argued for the advantages of conformity in many organizational contexts. They believe that companies stand to benefit by following typical repertoires, as these are perceived as attractive and legitimate by customers, competitors, and powerful stakeholders alike (Meyer & Rowan 1977; Scott & Meyer 1983). Thus, although we have argued that poor performance will promote nonconformity, we also expect that nonconformity itself will tend to generate decreases in performance.

Hypothesis 7. Competitive nonconformity will erode financial performance.

Certainly hypothesis 7 is a very general one that demands some specification of contexts. Competitive deviation is less likely to erode performance in large organizations that can much more easily support or legitimize their unorthodox activities (Haveman 1993b; Scott & Meyer 1987). Big firms have the visibility and market presence to generate acceptance for their new approach, the power to lobby government agencies, and the influence to ally themselves with partners who can help them indoctrinate customers (Mezias 1990; Miller 1990). Hence some giants can successfully resist — and ultimately transform — industry norms and conventions. But smaller organizations may be too poor and powerless to diverge from the pack (Haveman 1993a). Their relative obscurity and lack of resources make it less likely that their nonconformity will attract profitable trade.

Hypothesis 8. Competitive nonconformity is more apt to erode performance in small than in large firms. (That is, size will moderate the relationship between deviation and change in performance.)

Another potential moderator of the relationship between nonconformity and change in performance is the diversity of markets a firm faces. Target markets that are diverse will usually include at least some niches whose customers have unusual demands and whose competitors pose unusual threats. In order to succeed in these niches, a firm often will require an unorthodox approach (Scott 1987). Companies with diversified markets will also be able to spread their business risks: they can afford to gamble with unconventional methods in at least some of their markets, since they have others to fall back on (Chandler 1962; Palmer, Jennings & Zhou 1993).

Hypothesis 9. Competitive nonconformity is more apt to erode the performance of firms operating in homogeneous than in diverse markets. (That is, market diversity will moderate the relationship between deviation and change in performance.)

Method

SAMPLE

The scope of our initial database and our final sample was determined by our notions of markets and competition. We view markets as social structures among specific cliques of firms whose repertoires are based in part on the mutual observation of concrete actions (see Burt 1987; Granovetter 1985; White 1981). We have argued that competitive repertoires are very much shaped by the regular and continual interaction among rivals and their clients. Thus our sample consisted of enterprises of comparable market scope that were direct competitors of one another over a significant interval.

Our initial database consisted of all 32 "major" and "national" U.S. domestic airlines, excluding all regional airlines. The airline industry was chosen because of its competitiveness and its rich diversity of competitive tactics (Chen, Smith & Grimm 1992). The industry also contains firms about whose decisions there is abundant public information and whose business is concentrated in a single industry.

Unfortunately, 14 of the 32 airlines were niche rather than truly major carriers, and considerable information was missing for all these firms. More important, while the 18 "major" airlines competed and interacted with one another across the same competitive domain and over a common interval, the 14 smaller and more focused carriers did not. The 18 majors during 1979 to 1986 viewed each other as principal competitors; but they were not much influenced by the 14 nonmajors that were of far narrower market scope and more tied to and influenced by local conditions (such distinct local markets alone might account for any deviation from industry norms). Moreover, the potential competitive repertoires of the smaller airlines are constrained simply by the very scope and scale of their operations. Our 18 majors collectively accounted for more than 84% of the annual revenue passenger miles flown in the U.S. from 1979 to 1986 (versus less than 8% for the 14 airlines we dropped).

The period we chose to study was the postderegulation (but premerger) era, a time during which airlines would have to discover the rules of the game in order to compete effectively (Levine 1987). Such an interval was expected to provide much variety in competitive deviation both between airlines and over time.

For practical purposes, it is impossible to examine every action taken at all levels inside an organization. We therefore assess competitive deviation only in implemented, public, market-oriented actions — in those significant concrete actions that are observable by customers, competitors, and other industry participants. Minute adjustments such as isolated price changes were too incremental and invisible to fall within the scope of this study.

This research selected all competitive actions reported between 1979 and 1986 in *Aviation Daily*. Methods of "structured content analysis" were used to identify and classify these moves (see Harrigan 1980; Jauch, Osborn & Martin 1980; Miller & Friesen 1984). *Aviation Daily* is widely known in the industry as the publication that provides by far the most comprehensive information on airline strategy (Chen & MacMillan 1992; Chen, Smith & Grimm 1992).

IDENTIFICATION OF ACTIONS

An extensive eight-year review of every issue of *Aviation Daily* was conducted to document 21 types of competitive moves in the following categories: changes in price, promotional activities, product line or service changes, distribution channel alterations, market expansions, intraindustry mergers and acquisitions, and strategic alliances. This focus was derived from the strategic content research of Hatten, Schendel & Cooper (1978), Khandwalla (1981), Porter (1980, 1985), and Scherer (1980). It incorporated all actions reported in *Aviation Daily* that airlines took in these categories. The types include all the important competitive moves identified by Levine (1987) in his comprehensive review of airline competition since deregulation. The only omission from his list was the use of computerized reservation systems. This action was unreliably reported in our data and relates mostly to internal operating practices that were beyond the scope of our analysis. The total sample consisted of 891 actions. Our efforts to establish the comprehensiveness and reliability of the information reported in *Aviation Daily* are described in Chen, Smith & Grimm (1992) and in Miller & Chen (1994).

OPERATIONALIZATION OF VARIABLES

Competitive Nonconformity

The measure of competitive nonconformity reflects a firm's deviation from the annual industry norm in its repertoire of actions. This index, computed for each of the $j = 1, \ldots, 18$ airlines, for each of the $t = 1, \ldots, 7$ years, is based on the number of each of the $i = 1, \ldots, 21$ types of actions that a firm makes compared to the sample average for that year. Some kinds of actions (e.g., alliances) are much less common than others (e.g., price cuts). Thus to avoid underweighting the former vis-à-vis the latter we computed standardized values ($M = 0$, $SD = 1$) for each type of decision across all firms. We did this separately for each year to allow for annual differences in industry strategies. Then we took the absolute value of the standard scores $|z^{t}_{i,j}|$ to represent deviations from that year's industry average for each type of decision. These absolute values were summed across all 21 types of actions to yield an unadjusted cumulative index of deviation from the industry norm (Cronbach's $\alpha = 0.82$). The natural log of this sum was taken to normalize the coefficient.

$$\textit{Unadjusted deviation index}\,(t) = \ln\left(\sum_{i} |z^{t}_{i,j}|\right)$$

This index is not purely a measure of deviation, for it tends to increase with the number of actions taken by a firm, thereby producing tautological relationships with size, diversity, and even age. It was necessary to remove this effect by eliminating that portion of the measure that reflects only the amount of activity. To do this, the unadjusted index of deviation was regressed against a total annual index of competitive activity that we calculated for each firm.[3] Our final

adjusted index of competitive deviation is composed of the residuals obtained from the above regression and reflects only nonconformity.

In principle, it would have been possible to use the unadjusted deviation index in our analyses and to control for the amount of activity by employing it as an independent variable. In fact, however, problems of multicollinearity resulted from the strong relationships between the activity index and the other independent variables of size, diversity, and age.[4]

Deviation in Tactical versus Strategic Actions

To establish the range of application of our findings we computed deviation indexes both as described above for all 21 types of actions and for a subset of "tactical" actions only. It could be argued that nontactical actions occur so rarely that whenever they take place they imply some deviation from the industry norm. Therefore we wished to determine whether our findings would be replicated after we deleted nontactical actions. We define tactical actions as those that involve no significant or long-term investments in fixed assets, people, or structure and that are in general quite reversible (Chen, Smith & Grimm 1992). The 21 generic types of actions were classified as tactical or nontactical by three raters working independently. There was total agreement among the raters in making the classification.[5] Only seven kinds of actions were deemed to be nontactical: mergers and acquisitions, hub creations, feeder and interindustry alliances, new services, frequent flyer programs, and new airplane purchases.

Performance

Performance was measured by operating revenue per available seat mile flown — R/ASM. This index has the advantage of reflecting not only the load factor of an airline, that is, the percentage of seats it fills, but also the revenue produced thereby. In addition, it avoids counting as good performance high load factors that come at the cost of inadequate revenues. Revenue per available seat mile is also very appropriate for our analyses as it is updated for most airline executives very frequently and thus can have a rapid impact on strategy. Measures such as return on equity or return on assets were avoided, because they are influenced by interest rates, debt-equity ratios, tax anomalies, depreciation policy, and other factors having little connection with the actions composing our deviation measure. Profitability and cost indexes of performance are also less relevant, since the present study looks only at external, market-oriented actions, not internal operating decisions.

Organizational Variables

Firm age was measured in years and size was simply the natural log of the number of full-time employees. Slack resources were assessed using Bourgeois's (1981) measure of annual general and administrative expenses as a percentage of annual operating expenses obtained from the annual publication *Air Carrier Financial Statistics.*

Environmental Variables

Market diversity was measured by two indexes, the first measuring customer diversity, the second, competitor diversity. The diversity of customers was reflected by the distribution of a carrier's business across the 37 largest U.S. airports; this was computed as the Gibbs-Martin (Gibbs & Martin 1962) index of heterogeneity:

$$\sum_{K=1-37}\left[\frac{A's\ passengers\ at\ airport\ k}{A's\ passengers\ across\ 37\ airports}\right]^2$$

This index is an inverse index of diversity; it will equal 1 if the airline serves only one airport and will approach 0 if its passengers are equally divided across the airports. To make it reflect diversity, the sign of this index is reversed when we report our findings.

The diversity of competitors was assessed as follows:

$$\sum_{K=1-37}\left[\frac{A's\ passengers\ at\ airport\ k}{A's\ passengers\ across\ 37\ airports}\right] * N\ of\ rivals\ at\ k$$

This index will be large to the extent that airline A's important markets have many competitors.

Information was obtained from the U.S. Department of Transportation's publication titled *Airport Activity Statistics of Certified Route Carriers*. Its data provide some indication of the variety of markets and passengers served and hence of the diversity of competitive conditions and customer demands that airlines must confront (Levine 1987).

Market growth was measured by the level of growth in total industry business volume, an indicator of the degree to which new markets or niches are opening up (Levine 1987). This was assessed by the annual change in total revenue passenger miles flown for the entire domestic airline passenger business.

We chose as our measure of time the number of years since deregulation in October 1978. Deregulation was an environmental jolt that severely disrupted the old methods of competition and threw the industry into competitive disarray and chaos. As time passed, the industry again began to settle down and become more competitively isomorphic (as our analyses will show).

ANALYSES

Since data were collected on 18 airlines over 8 years it was useful to correct for potential autocorrelation due to time and for heteroskedasticity caused by cross sections (Sayrs 1989). To do this we employed Kmenta's (1986) autoregressive-heteroskedastic model for pooled time series — cross-sectional data. The Kmenta model is

$$Y_{it} = b_1X_{it,1} + b_2X_{it,2} + \ldots. + b_kX_{it,k} + E_{it}$$

where i equals 1, 2, . . . , N, t equals 1, 2, . . . , T, k equals 1, 2, . . . , K. K equals the number of explanatory variables, N equals the number of cross-sectional units (18), and T equals the number of time periods (7 — one was lost due to

TABLE 2A: Descriptive Statistics

	Mean (Deviation)	Standard Deviation Indexes
Competitive nonconformity		
All actions (log)	.017	.260
Tactical actions (log)	.003	.218
Organization		
Size (log)	9.240	1.229
Slack resources	.112	.043
Firm age	30.262	17.140
Performance	.083	.016
Environment		
Market growth	3.96	7.85
Customer diversity	.213	.109
Competitor diversity	12.27	2.94
Time since deregulation	4	2

lagging). We applied the above model to data corrected for autocorrelation and heteroskedasticity. To correct for first-order autocorrelation, the Prais-Winsten transformation, which preserves the first observation of each cross section, was applied to each firm individually. Plots of residuals, rapidly declining correlograms, and nonsignificant Durbin *h* statistics revealed that all autocorrelation had been removed by the correction and that no higher-order autocorrelation was present in the data. In a second step, the data were corrected for heteroskedasticity by dividing dependent and independent variables by the firm-specific error variance obtained from the regressions on the autocorrelation-corrected data. Subsequent plots of residuals revealed an absence of heteroskedasticity. The twice-transformed data could now be pooled and analyzed using OLS regression (see also Judge et al. 1988, Section 11.5).

In examining the moderating effects of size and diversity on changes in performance (Table 4), the components of the interaction terms were standardized before multiplication to remove any multicollinearity with the main effects (Smith & Sasaki 1979; Southwood 1978). Separate models were run for each interaction analysis to avoid multicollinearity between the interaction terms. For each model, the significance levels reported are those computed for the additional variance explained by the interactions after all noninteraction variables have been introduced into the model.

TABLE 2B: Pearson Correlation Matrix

	Nonconformity All Actions	Nonconformity Tactical Actions	Firm Size	Slack Resources	Firm Age
All actions	1.00				
Tactical actions	.78	1.00			
Firm size	.33	.16	1.00		
Slack resources	.02	-.01	-.23	1.00	
Firm age	.18	.03	.61	-.24	1.00
Performance	-.24	-.15	-.34	.04	-.55
Market growth	.22	.06	-.03	.03	.11
Customer diversity	.29	.18	.56	-.20	.27
Competitor diversity	.26	.11	.44	.15	.58
Time since deregulation	.16	-.02	-.07	-.02	.12

	Performance	Market Growth	Customer Diversity	Competitor Diversity
Performance	1.00			
Market growth	-.13	1.00		
Customer diversity	-.20	.01	1.00	
Competitor diversity	-.27	.49	.25	1.00
Time since deregulation	-.16	.59	-.00	.55

[a] For the smallest sample size, all coefficients greater than .19 and .24 are significant at beyond the 0.05 and 0.01 levels, respectively.

Findings

Descriptive statistics and correlations are presented in Table 2, and the results of the multiple regression analyses are given in Tables 3 and 4.

Table 3 indicates that hypothesis 1 received partial support while hypothesis 2 was well supported: the market context did influence the level of nonconformity. Operating in markets with a diversity of customers was associated with overall nonconformity, but not with tactical nonconformity. By contrast, having a diversity of competitors contributed to tactical but not to overall nonconformity. Perhaps a diversity of customers and markets generates ideas and demands for creative combinations of new services, hubs, alliances, client programs, and equipment alterations — the kinds of actions that we deemed to be nontactical. But day-to-day competition with a diversity of rivals appears to elicit noncon-

TABLE 3: Regressions of Competitive Nonconformity on Airline and Industry Characteristics

	Standardized Coef.	
	All Actions	Tactical Actions
Customer diversity	.11*	.01
Competitor diversity	.07	.23*
Time since deregulation	-.34**	-.55***
Firm size	.20*	.12
Slack resources	.11+	.08
Firm age	-.22*	-.34**
Revenue/available seat mile	-.26***	-.26**
Market growth	.50***	.40**
R^2	.33	.15
F ratio	7.20	2.61
Degrees of freedom 8, 118		

Significance for partial F statistics:
+ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

formity in the profiles of more routine actions such as price increases and decreases, route entries and exits, and small promotions — actions that are the stuff of everyday rivalry.

As predicted by hypothesis 2, nonconformity declined with the time since deregulation. After the shock and some initial groping and experimentation, companies began to resemble each other more and more in the way they competed (see DiMaggio & Powell 1983, 1988).

There was mixed support for the firm-level variables of our model. We found modest relationships of size and slack resources to total nonconformity but not to tactical nonconformity. Hence hypothesis 3 was not well substantiated, perhaps because we focused only on the largest airlines. There was support, though, for hypothesis 4: older firms were indeed more reluctant to deviate from industry norms.

Hypotheses 5 and 6 were supported as well. Both a climate of market growth and poor performance seemed to provoke competitive nonconformity for the total set of actions and for the tactical set of actions. The economic component of our model was strongly borne out.

The results in Table 4 confirm that competitive nonconformity was indeed associated with declines in performance (hypothesis 7). The central tendencies in an industry tend to be somewhat functional, and deviations from them may

TABLE 4A: Regressions of Change in Performance on Nonconformity, Airline, and Industry Characteristics: Size and Customer Diversity Interaction

	Standardized Coefficients			
	Size Interaction		Customer Diversity Interaction	
	All Actions	Tactical Actions	All Actions	Tactical Actions
Competitive nonconformity	-.20***	-.20***	-.19***	-.16***
Customer diversity	.11*	.12*	.19**	.14*
Competitor diversity	-.16*	-.17*	-.15*	-.13*
Time since deregulation	-.31***	-.32***	-.34***	-.35***
Firm size	.01	-.02	-.05	-.07
Slack resources	.03	.03	.02	.03
Firm age	.06	.06	.09	.08
Market growth	.07	.05	.10	.06
Deviation * size	.20***	.17***		
Deviation * customer diversity (-)			.22***	.15**
R^2	.53	.52	.51	.50
F ratio	14.51***	14.17***	13.58***	13.81***
Degrees of freedom 9, 117				

be hazardous. Although some deviations no doubt pay off, it seems that in our sample, the risks and dangers outweigh the rewards. And this is consistent with the views of many institutionalists (Meyer & Rowan 1977; Zucker 1987).

The results in Table 4, however, suggest caution in evaluating hypotheses 8 and 9. Size and customer diversity were important moderators of the relationship between nonconformity and change in performance. Specifically, the relationship between nonconformity and performance became positive for the largest firms and for those with the most diverse markets. In order to assess the interaction effects more precisely, we calculated saddle points for the regression equations using the method of partial derivatives outlined by Gupta & Govindarajan (1993) and Harrison & Mitchell (1995). Only the largest 15% or so of our firms show a positive relationship between nonconformity and performance, while 85% show a negative relationship. The results are essentially the same for tactical nonconformity, with the percentages being 86% and 14% respectively. For customer diversity, only those 21% of firms with the greatest

TABLE 4B: Regressions of Change in Performance on Nonconformity, Airline, and Industry Characteristics: Competitor Diversity Interaction

	Standardized Coef.	
	All Actions	Tactical Actions
Competitive nonconformity	-.09*	-.12**
Customer diversity	.07	.09+
Competitor diversity	-.16*	-.13*
Time since deregulation	-.27**	-.35***
Firm size	.01	-.03
Slack resources	.03	.04
Firm age	.06	.06
Market growth	-.03	-.00
Deviation * competitor diversity	.07+	.02
R^2	.49	.49
F ratio	11.35	12.36
Degrees of freedom 9, 117		

Significance for partial F statistics:

+ p < .10 * p < .05 ** p < .01 *** p < .001

diversity exhibited a positive relationship between nonconformity and performance; the rest had a negative relationship. For tactical nonconformity, the corresponding percentage is also 14%. The interaction effect for competitor diversity, however, was almost nonsignificant, perhaps because such diversity is less indicative of a firm's need to cater to different market demands. To recap, nonconformity is useful only for the top quintile or so of airlines with the largest size and the greatest customer diversity.

It was encouraging that most of the hypotheses were supported and that the results converged both for the whole set of decisions and for the tactical decisions alone. It is also notable that findings based on a seven-year rather than an annual standard for assessing nonconformity (available from the first author) replicated most of the results of the present study.

Discussion and Conclusion

SOCIOLOGICAL APPROACHES TO COMPETITIVE NONCONFORMITY

Competitive nonconformity is sometimes attributed to the genius or aberrancy of top managers or the corporate cultures these people create (Miller 1990). And when considered by economists, nonconformity is viewed in the context of distinctive competencies and their economic rents (Porter 1980, 1985; Prahalad & Hamel 1990). No doubt these are valuable approaches. One of the most interesting findings to emerge from this study, however, is that nonconformity also can be studied froma socIt is no contradiction to the value of these findings to note that markets can be viewed as social structures that over time exert a socializing influence on their members (Granovetter 1985; White 1981). As Wellman and Berkowitz (1988) state, "markets are social structures and not the spontaneous products of aggregated dyadic exchanges. . . . The ways in which actors observe and respond to others — the essence of a market — can be determined in the same fashion as any other socially structured activity" (222). Indeed, we found that the nonconformity of market participants arose in part from their social context and in part from their relationships to that context. Specifically, the extensiveness and continuity of a firm's interactions with its market contributed significantly to its level of nonconformity.

But we do not embrace a deterministic or "oversocialized" view of markets. A firm's temporal and extensive degree of embeddedness within the existing structure of social relations is of crucial importance, and this varies both among contexts and among firms within a context. Hence we must reject both the determinism of universal norms and the atomistic rationality of economists. Perhaps Granovetter (1985) says it best: "Actors do not behave or decide as atoms outside a social context, nor do they adhere slavishly to a script written for them by the particular intersection of social categories that they happen to occupy. Their attempts at purposive action are instead embedded in concrete, ongoing systems of social relations" (487). Indeed, concrete "social relations, rather than institutional arrangements" are most essential, and these "networks of social relations penetrate irregularly and in differing degrees" even in the same market (Gronovetter 1985:491).

The current findings, tentative though they may be, provide early empirical confirmation of the usefulness of a sociological approach to markets (Granovetter 1985; White 1981, 1988; Winship & Rosen 1988). The sociological approach has, it seems, much to say about how competitive repertoires are shaped, not just in regulated contexts, but in highly competitive markets. Our research also shows how nonconformity can be operationalized objectively and as a continuous variable. This may enable subsequent researchers to avoid "all or nothing," and hence oversocialized and undersocialized, market perspectives.

Our enthusiasm for the sociological view notwithstanding, we should not forget that nonconformity is influenced powerfully by economic stimuli as well — most importantly by the rate of market growth and poor financial returns. Sociological perspectives complement but do not replace economic approaches to the study of competition (Powell 1991; Winship & Rosen 1988; Zald 1978). Although the two approaches stress different sources of nonconformity, each is

important, and together they provide much incentive to broaden the conceptual rationales used to study markets.

Neither must we overstate the separation between economic and sociological influences. As one perceptive reviewer has suggested, the economic and sociological perspectives of markets vary in the observables that are taken to determine behavior. Economists concentrate more heavily on prices and costs, whereas sociologists, who also recognize the importance of these factors, are interested as well in the social contexts that influence which observables managers see and react to. This set of stimuli can be quite large. And as we indicated in the discussions preceding hypotheses 5 and 6, economic incentives may be influenced by the social context. Similarly, some types of social interaction might evoke economic reasoning. It may be as Braudel (1982) claims, that "all forms of exchange are both economic and social" (227).

NONCONFORMITY AND PERFORMANCE

Many institutionalists see organizations as adopting various stratagems, policies, and administrative practices not so much to improve their efficiency or reliability but rather to secure the legitimacy that will gain access to resources and status (Meyer & Rowan 1977). According to DiMaggio & Powell (1983) a primary mechanism that organizations use to achieve legitimacy is conformity to industry norms. Deviant organizations that use unorthodox techniques to compete may be punished (Chen & Hambrick 1995). This, to some degree, is borne out by our results — but mostly for smaller firms operating in homogeneous markets. One reason for the superior results of companies pursuing "isomorphic" or conventional repertoires may be that they are more acceptable to competitors, customers, and even suppliers of capital. It may also be that deviant repertoires are risky in and of themselves — they ignore the wisdom that is often reflected in industry consensus.

The arguments of the institutionalists do not appear to apply, however, to large and diverse organizations that can actually benefit from their nonconformity. This exception is important, for it is precisely in industries with few large firms where institutional theory should receive the greatest support. Mezias (1990) and Scott (1987) seem to be right, then, in calling for the qualification and contextualization of institutional hypotheses.

These results have potential implications for industries such as airlines that have been undergoing a rash of mergers. Initially, the creation of organizations that are free of tradition may give rise to unconventional strategies. As we have seen, new companies and industry upheavals make for a wider range of idiosyncratic strategies. Ultimately, however, mergers cause industry concentration, and this may reduce diversity and thus enhance the conformity of strategies. In this regard, the U.S. automobile industry may be a suitable target for study.

This research also has implications for scholars of strategic management. Barney (1991), Porter (1980, 1985), Prahalad and Hamel (1990), and others have argued for the benefits of developing distinctive competencies. To the extent that these competencies demand unique competitive profiles, however, there may be some constraints on this pursuit, especially for small firms operating in

homogeneous markets. For them, the development of distinctive competencies should probably not come at the expense of abandoning important industry conventions. In these instances, competitive conformity may have more advantages than deviation in procuring resources and legitimacy, meeting the expectations of customers, regulators, and other stakeholders, and avoiding hostile retaliation.

Clearly, this research raises more questions than it answers. The notion of competitive nonconformity is new and could benefit from more precise articulation and alternative definition. Another limitation of this study is that we could not discover the precise motivations for nonconformity, nor could we assess the performance implications of the individual components of nonconformity. Moreover, internal policies and resources have much to do with competitive deviation and originality (Barney 1991), yet these could not be examined in this research. Company structures, cultures, and top management teams also will no doubt influence nonconformity, but these were not treated here. (For one investigation, see Keck & Tushman 1993.) And there are other institutional forces such as legal, professional, and competitive contexts that might have a bearing on competitive conformity (DiMaggio & Powell 1991). All these gaps in this study suggest avenues of further research as does the challenge of replicating our findings in other industries.

Notes

1. We would expect that periods of precipitous decline or collapse in an industry would also trigger deviations from the norm as companies seek new ways to survive. But this assymetrical relation could not be tested, since our industry never did experience such an interval (see also Lieberson 1985:124).

2. Some might argue that poor performance erodes organizational resources, and that this increases the pressure to conform. This usually will be true only of long-term performance deterioration. Indeed, resource scarcity is the concern of hypothesis 6.

3. The activity index was composed of the same elements as the index of deviation but without taking their absolute value:

$$\textit{Activity index}(t) = \ln\left(\sum_{I} z_{i,j}^{t}\right)$$

4. We also assessed the possibility of deriving homogeneous groups or clusters of airlines based on repertoire similarity and then using distinct standards for assessing nonconformity within each group. Unfortunately, no clear groups emerged from various clustering attempts (single and average linkage algorithms using Euclidean distance measures and Ward's criterion) or from *Q*-factor analysis.

Clustering was performed using the standardized scores for the 21 types of actions. We found that airlines moved between clusters from year to year and that clusters seemed to vary only slightly and in no thematic way — the meanings of most clusters were obscure. Moreover, one cluster accounted for most observations. The small "fringe" clusters that emerged contained mostly regional carriers that were ultimately deleted from the sample in this study. When we used a holdout sample to establish the stability of the clusters, the original clusters were not replicated. Matters did not improve after we had removed the "outlier" clusters, which merely produced a few new outliers. (It seemed inadvisable to use the small clusters as a basis for calculating deviations, as these were at odds with industry norms, and their members merely conformed to the behavior of a similarly unorthodox partner.)

We also found it impractical to derive clusters based on geography or on the frequency of competitive interaction among the airlines. Virtually all the larger airlines interacted with each other in a great many of their routes. A few of the smaller airlines did emerge as being outside

the pack, but these represented fewer than 6% of our observations. Given these results, it seemed reasonable to use industry central tendencies as the bases for assessing nonconformity.

5. To confirm the validity of the classification, a questionnaire was distributed to six management professors with expertise in both competitive strategy and the airline industry. The respondents were asked to categorize various types of actions as strategic or as tactical using the above definitions. Interrater reliability was 1.0 for 20 items, and 0.83 for one item. There was only a single misclassification in the entire exercise.

References

Aldrich, Howard E., and Ellen Auster. 1986. "Even Dwarfs Started Small: Liabilities of Size and Age and Their Strategic Implications." Pp. 165-98 in *Research in Organizational Behavior*. Vol. 8, edited by Barry Staw and Larry L. Cummings. JAI Press.

Amburgey, Terry, and Ann Miner. 1992. "Strategic Momentum: The Effects of Repetitive, Positional and Contextual Momentum on Merger Activity." *Strategic Management Journal* 13:335-48.

Barney, Jay. 1991. "Firm Resources and Sustained Competitive Advantage." *Journal of Management* 17:99-120.

Baum, Joel A.C., and Christine Oliver. 1992. "Institutional Embeddedness and the Dynamics of Organizational Populations." *American Sociological Review* 57:540-59.

Berger, Peter, and Thomas Luckmann. 1967. *The Social Construction of Reality*. Doubleday.

Blau, Peter M., and Richard Schoenherr. 1971. *The Structure of Organizations*. Wiley.

Boeker, Warren. 1989. "Strategic Change: The Effects of Founding and History." *Academy of Management Journal* 32:489-515.

Bourgeois, L. Jay. 1981. "On the Measurement of Organizational Slack." *Academy of Management Review* 6:29-39.

Braudel, Fernand. 1982. *Les jeux de l'échange*. A. Colin.

Burt, Ronald S. 1987. "Social Contagion and Innovation: Cohesion versus Structural Equivalence." *American Journal of Sociology* 92:1287-1335.

Bylinsky, Gene. 1974. *The Innovation Millionaires*. Knopf.

Chandler, Alfred D. 1962. *Strategy and Structure*. MIT Press.

____. 1990. *Organizational Scale and Scope*. Wiley.

Chen, Ming-Jer. 1988. "Competitive Strategic Interaction: A Study of Competitive Actions and Responses." Unpublished Ph.D. diss., University of Maryland.

Chen, Ming-Jer, and Donald C. Hambrick. 1995. "Speed, Stealth and Selective Attack: How Small Firms Differ from Large Firms in Competitive Behavior." *Academy of Management Journal* 38:453-82.

Chen, Ming-Jer, and Ian MacMillan. 1992. "Nonresponse and Delayed Response to Competitive Moves: The Roles of Competitor Dependence and Irreversibility." *Academy of Management Journal* 35:539-70.

Chen, Ming-Jer, and Danny Miller. 1994. "Competitive Attack, Retaliation and Performance." *Strategic Management Journal* 15:85-102.

Chen, Ming-Jer, Ken G. Smith, and Curtis Grimm. 1992. "Action Characteristics as Predictors of Competitive Responses." *Management Science* 38:439-55.

Cohen, William, and Daniel Levinthal. 1990. "Absorptive Capacity: A New Perspective on Learning and Innovation." *Administrative Science Quarterly* 35:128-52.

Cooper, Arnold C., Gary Willard, and Carolyn Woo. 1986. "Strategies of High Performing New and Small Firms: A Re-examination of the Niche Concept." *Journal of Business Venturing* 1:247-60.

Cyert, Richard M., and James G. March. 1963. *A Behavioral Theory of the Firm*. Prentice-Hall.

DiMaggio, Paul J. 1991. "Constructing an Organizational Field as a Professional Project: U.S. Art Museums, 1920-1940." Pp. 267-92 in *The New Institutionalism in Organizational Analysis*, edited by Walter W. Powell and Paul DiMaggio. University of Chicago Press.

DiMaggio, Paul J., and Walter W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48:147-60.

____. 1988. "Interest and Agency in Institutional Theory." Pp. 3-21 in *Institutional Patterns and Organizations*, edited by Lynne Zucker. Ballinger.

____. 1991. "Introduction." Pp. 1-40 in *The New Institutionalism in Organizational Analysis*, edited by Walter W. Powell and Paul J. DiMaggio. University of Chicago Press.

Eisenhardt, Kathleen. 1989. "Making Fast Strategic Decisions in High-Velocity Environments." *Academy of Management Journal* 32:543-76.

Erickson, B. 1988. "The Relational Basis of Attitudes." Pp. 99-121 in *Social Structures*, edited by Barry Wellman and Stanley. Berkowitz. Cambridge University Press.

Fligstein, Neil. 1985. "The Spread of the Multidivisional Form among Large Firms: 1919-1979." *American Sociological Review* 50:377-91.

Gibbs, Jack P., and Walter T. Martin. 1962. "Urbanization, Technology and the Division of Labour." *American Sociological Review* 26:667-77.

Granovetter, Mark. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78:1360-80.

____. 1985. "Economic Action and Social Structure: The Problem of Embeddedness." *American Journal of Sociology* 91:481-510.

Gupta, Anil, and V.G. Govindarajan. 1993. "Methodological Issues in Testing Contingency Theories." Pp. 121-56 in *Creative and Innovative Approaches to the Science of Management*, edited by Yuri Ijiri. Quorum.

Halberstam, David. 1986. *The Reckoning*. Avon.

Hannan, Michael T., and John H. Freeman. 1984. "Structural Inertia and Organizational Change." *American Sociological Review* 49:149-64.

Harrigan, Kathryn. 1980. *Strategies for Declining Businesses*. Lexington.

Harrison, J. Richard, and Larry Mitchell. 1995. "Interaction Effects in Quasi-Linear Models." Working paper, University of Texas at Dallas.

Hatten, Kenneth J., Dan E. Schendel, and Arnold C. Cooper. 1978. "A Strategic Model of the U.S. Brewing Industry: 1952-1971." *Academy of Management Journal* 22:592-610.

Haveman, Heather A. 1993a. "Organizational Size and Change: Diversification in the Savings and Loan Industry after Deregulation." *Administrative Science Quarterly* 38:20-50.

____. 1993b. "Follow the Leader: Mimetic Isomorphism and Entry into New Markets." *Administrative Science Quarterly* 38:593-627.

Homans, George C. 1950. *The Human Group*. Harcourt, Brace.

Jacobs, David, Michael Useem, and Mayer Zald. 1991. "Firms, Industries, and Politics." *Research in Political Sociology* 5:141-65.

Jauch, Lawrence R., Robert Osborn, and Thomas Martin. 1980. "Structured Content Analysis of Cases: A Complementary Method for Organizational Research." *Academy of Management Review* 5:517-26.

Judge, George, R. Hill, W. Griffiths, H. Lutkepohl, and T. Lee. 1988. *Introduction to the Theory and Practice of Econometrics*. 2d ed. Wiley.

Keck, Sara, and Michael L. Tushman. 1993. "Environmental and Organizational Context and Executive Team Structure." *Academy of Management Journal* 36:1314-44.

Khandwalla, Pradip N. 1981. "Properties of Competing Organizations." Pp. 256-89 in *Handbook of Organizational Design*. Vol. 1, edited by Paul Nystrom and William H. Starbuck. Oxford University Press.

Kmenta, Jan. 1986. *Elements of Econometrics.* 2d ed. Macmillan.

Kreps, David. 1990. *Game Theory.* Oxford.

Lant, Theresa K., and David B. Montgomery. 1987. "Learning from Strategic Success and Failure." *Journal of Business Research* 15:503-18.

Levine, Michael E. 1987. "Airline Competition in Deregulated Markets: Theory, Firm Strategy and Public Policy." *Yale Journal on Regulation* 4:393-494.

Lieberman, Marvin, and David Montgomery. 1988. "First Mover Advantage." *Strategic Management Journal* 9:441-58.

Lieberson, Stanley. 1985. *Making It Count: The Improvement of Social Research and Theory.* University of California.

March, James G., and Herbert Simon. 1958. *Organizations.* Wiley.

Meyer, John W., and Brian Rowan. 1977. "Institutionalized Organizations: Formal Structure as Myth and Ceremony." *American Journal of Sociology* 82:340-63.

Meyer, Marshall W., and Lynne Zucker. 1989. *Permanently Failing Organizations.* Sage.

Mezias, Stephen J. 1990. "An Institutional Model of Organizational Practice: Financial Reporting at the Fortune 200." *Administrative Science Quarterly* 35:431-57.

Miller, Danny. 1990. *The Icarus Paradox.* HarperCollins.

____. 1993. "The Architecture of Simplicity." *Academy of Management Review* 18:116-38.

Miller, Danny, and Ming-Jer Chen. 1993. "The Simplicity of Strategic Repertoires." Pp. 32-37 in *Best Paper Proceedings of the Academy of Management.*

____. 1994. "The Causes and Consequences of Competitive Inertia." *Administrative Science Quarterly* 39:1-24.

Miller, Danny, and Peter H. Friesen. 1984. *Organizations: A Quantum View.* Prentice-Hall.

Milliken, Frances J., and Theresa K. Lant. 1991. "The Impact of an Organization's Recent Performance History on Strategic Persistence and Change." Pp. 129-56 in *Advances in Strategic Management.* Vol. 7, edited by Paul Shrivastava, Anne Huff, and Jane Dutton. JAI Press.

Oliver, Christine. 1988. "The Collective Strategy Framework: An Application to Competing Predictions of Isomorphism." *Administrative Science Quarterly* 33:543-61.

Palmer, Donald A., P. Devereaux Jennings, and X. Zhou. 1993. "Late Adoption of the Multidivisional Form by Large U.S. Corporations: Institutional, Political and Economic Accounts." *Administrative Science Quarterly* 38:100-31.

Pfeffer, Jeffrey. 1981. *Power in Organizations.* Ballinger.

Pfeffer, Jeffrey, and Gerald Salancik. 1978. *The External Control of Organizations.* Harper & Row.

Porter, Michael E. 1980. *Competitive Strategy.* Free Press.

____. 1985. *Competitive Advantage.* Free Press.

____. 1991. "Toward a Dynamic Theory of Strategy." *Strategic Management Journal* 12:95-118.

Powell, Walter W. 1985. *Getting into Print: The Decision Making Process in Scholarly Publishing.* University of Chicago Press.

____. 1991. "Expanding the Scope of Institutional Analysis." Pp. 183-203 in *The New Institutionalism in Organizational Analysis,* edited by Walter W. Powell and Paul J. DiMaggio. University of Chicago Press.

Prahalad, C.K., and Gary Hamel. 1990. "The Core Competence of the Corporation." *Harvard Business Review,* May-June, pp. 79-91.

Rumelt, Richard P. 1974. *Strategy, Structure and Economic Performance.* Harvard Business School.

Rumelt, Richard P., Dan Schendel, and David Teece. 1991. "Strategic Management and Economics." *Strategic Management Journal* 12:5-30.

Sayrs, Lois. 1989. *Pooled Time Series Analysis.* Sage.

Scherer, Fredrick M. 1980. *Industrial Market Structure and Economic Performance.* Houghton-Mifflin.

Schumpeter, Joseph A. 1950. *Capitalism, Socialism and Democracy.* Harper & Row.

Scott, W. Richard. 1987. "The Adolescence of Institutional Theory." *Administrative Science Quarterly* 32:493-511.

Scott, W. Richard, and John W. Meyer. 1983. "The Organization of Societal Sectors." Pp. 230-67 in *Organizational Environments: Ritual and Rationality,* edited by John W. Meyer and W. Richard Scott. Sage.

Scott, W. Richard, and John W. Meyer. 1987. "Environmental Linkages and Organizational Complexity." Pp. 129-53 in *Comparing Public and Private Schools,* edited by H. Levin and T. James. Falmer Press.

Shamsie, Jamal. 1989. "First Mover Advantage." Working paper, Faculty of Management, McGill University.

Singh, Jitendra V. 1990. *Organizational Evolution.* Sage.

Smith, Ken W., and M. Sasaki. 1979. "Decreasing Multicollinearity: A Method for Models with Multiplicative Functions." *Sociological Methods and Research* 8,1:35-56.

Smith, Ken G., Curtis Grimm, Mertin Gannon, and Ming-Jer Chen. 1991. "Organizational Information Processing, Competitive Responses and Performance in the U.S. Domestic Airline Industry." *Academy of Management Journal* 34:60-85.

Southwood, Kenneth E. 1978. "Substantive Theory and Statistical Interaction: Five Models." *American Journal of Sociology* 83:1154-1203.

Spender, J.C. 1989. *Industry Recipes: The Nature and Source of Managerial Judgement.* Blackwell.

Starbuck, William H. 1983. "Organizations as Action Generators." *American Sociological Review* 48:91-102.

Starbuck, William H., Arendt Greve, and Bo Hedberg. 1978. "Responding to Crises." *Journal of Business Administration* 9:111-37.

Starbuck, William H., and Frances J. Milliken. 1988. "Challenger: Fine-Tuning the Odds Until Something Breaks." *Journal of Management Studies* 25:319-40.

Staw, Barry M. 1981. "Escalation of Commitment to a Course of Action." *Academy of Management Review* 6:577-87.

Stinchcombe, Arthur L. 1965. "Social Structure and Organizations." Pp. 142-92 in *Handbook of Organizations,* edited by James G. March. Rand McNally.

Tushman, Michael L., and Philip Anderson. 1988. "Technological Continuities and Organizational Environments." *Administrative Science Quarterly* 33:439-65.

Tushman, Michael L., and Elaine Romanelli. 1985. "Organizational Evolution: A Metamorphosis Model of Convergence and Reorientation." Pp. 171-222 in *Research in Organizational Behavior.* Vol. 7, edited by Larry L. Cummings and Barry M. Staw. JAI Press.

Tversky, Amos, and Daniel Kahneman. 1973. "Availability." *Cognitive Psychology* 5:207-32.

Wellman, Barry, and Stanley Berkowitz. 1988. *Social Structures.* Cambridge University Press.

Wernerfelt, Birger, and Aneel Karnani. 1987. "Competitive Strategy under Uncertainty." *Strategic Management Journal* 8:187-94.

White, Harrison C. 1981. "Where Do Markets Come From?" *American Journal of Sociology* 87:517-47.

____. 1988. "Varieties of Markets." Pp. 226-57 in *Social Structure,* edited by B. Wellman and S. Berkowitz. Cambridge University Press.

White, Kenneth J. 1993. *SHAZAM.* Econometrics Computer Program. McGraw-Hill.

Winship, Charles, and Stuart Rosen. 1988. "Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure." *American Journal of Sociology* 94:S1-S16.

Zald, Mayer. 1978. "The Social Control of Industries." *Social Forces* 57:79-102.

Zelditch, Morris, Jr., and Henry A. Walker. 1984. "Legitimacy and the Stability of Authority." Pp. 65-90 in *Advances in Group Processes*, edited by Edward E. Lawler. JAI Press.

Zucker, Lynne. 1987. "Institutional Theories of Organization." *Annual Review of Sociology* 13:443-64.

Labor Market Structures in Japan: An Analysis of Organizational and Occupational Mobility Patterns*

MARIAH MANTSUN CHENG, *University of North Carolina at Chapel Hill*
ARNE L. KALLEBERG, *University of North Carolina at Chapel Hill*

Abstract

In this article, we examine patterns of occupational and organizational mobility among Japanese men. Contrary to conventional wisdom, we find that Japanese men often change employers as well as occupations during their careers. Using a conceptualization of internal labor markets that distinguishes between firm internal labor markets (FILMs) and occupational internal labor markets (OILMs), we tested various hypotheses about the determinants of four patterns of job shifts formed by cross-classifying firm and occupational mobility. Our results suggest that men in larger organizations (who we assumed are more apt to be in FILMs) are more likely than those in smaller firms to experience within-firm mobility, while men in tiny firms make more across-firm shifts. In addition, men in professional and technical occupations (who we assume are more likely to be in OILMs) experience more within-occupation mobility.

Japanese workers are commonly perceived as having relatively stable relationships with their employers. Publicity about the cultural uniqueness of permanent employment and managerial styles in Japan (e.g., Abegglen 1958; Vogel 1979) has helped to foster the belief that Japanese employees work for the same employer almost all their lives and enjoy numerous internal promotion opportunities. Academic discussions also regard the lifelong employment and seniority wage system as two of the "three pillars" of Japanese industrial relations (the third being enterprise unionism) that have facilitated the nation's

* *Revision of a paper presented at the 1994 annual meeting of the American Sociological Association in Los Angeles. We are grateful to Aage B. Sørensen, Robert P. Althauser, Michael Hout, David Guilkey, Kazuo Yamaguchi, Rachel Rosenfeld, Arthur Sakamoto, Guang Guo, Lynn Igoe, and the anonymous reviewers for their helpful comments at various points as we developed this article. Thanks are also due to the Carolina Population Center, University of North Carolina at Chapel Hill, because this research was made possible by its support through a postdoctoral fellowship funded by the NICHD National Research Service Award (5-T32-HD07168-16). Direct correspondence to Mariah Mantsun Cheng, Carolina Population Center, University of North Carolina, Chapel Hill, NC 27516-3997.*

immense economic growth since World War II (Odaka 1986). It is widely believed that Japan's path to modernization not only led to industrial excellence, but also provided its workforce with employment security and seniority-based opportunities for advancement.

However, a growing body of evidence suggests that this formulaic presentation of Japanese employment relations is misleading (Plath 1983; Shirai 1983). Although job mobility is lower in Japan than in the U.S. (Lincoln & Kalleberg 1990; Mincer & Higuchi 1988; Tachibanaki 1984), studies have found a considerable amount of career mobility among Japanese men. Levine (1983) calculated that, on average, a Japanese man changes firms at least three times during his work life, while Hashimoto and Raisian (1985) found that a Japanese man typically changes jobs five times before he retires at age 65. Thus, the notion of "lifetime" employment undoubtedly overstates the stability of the Japanese employment system and underestimates actual job mobility.

Dwelling on the cultural uniqueness of Japan's system overlooks variations in employment experiences among different sociodemographic groups in Japan and neglects to consider that many "unique" Japanese phenomena can be accounted for by universal processes common to industrial societies (Aoki 1984; Cole 1979; Levine 1983; Taira 1970). Many scholars point out that lifetime employment applies only to regular primarily male employees in large, private firms or government agencies (Dore 1973; Marsh & Mannari 1971; Taira 1962).[1] Young women, though sometimes hired as regular employees in large firms, are expected to leave after marriage or first childbirth; older women who rejoin the labor force commonly do so as temporary workers (Brinton, Ngo & Shibuya 1991; Rohlen 1979). Consequently, women are customarily excluded from the lifetime employment system, as are men in small companies, and temporary and "outside workers" in large firms (Chalmers 1989). Some researchers even suggest that employees under a lifetime system constitute no more than a fifth of Japan's work force (Cole 1979; Taira 1962).

Unfortunately, despite widespread interest in Japan's employment system, there is little empirical evidence on patterns of employer changes. We provide evidence on one important aspect of this system: patterns of occupational and organizational mobility among Japanese men. We argue that the directions and kinds of job shifts reveal much about types of firm and occupational labor market structures. We first describe the structure of labor markets in Japan. We then generate eight hypotheses about the correlates of inter- and intra-occupational and firm mobility patterns in Japan, and test them by multinomial logistic regression analyses of retrospective mobility data from a Japanese national survey.

Structure of the Japanese Labor Market

INDUSTRIAL DUALISM AND DUAL LABOR MARKETS

Three related notions are frequently invoked in discussions of Japanese labor market structures: industrial dualism, dual labor markets, and firm internal labor markets (FILMs). Japan is described as having a dual economy polarized by firm size, internal organization, and wage structure. One sector (the "core")

consisting of large-scale, capital-intensive organizations and high-wage systems, coexists with a second sector (the "periphery") characterized by many small-to-medium scale, labor-intensive, and low-wage firms (Broadbridge 1966; Minami 1986; Nakamura 1981).[2] While a dualistic economic structure is certainly not unique to Japan (e.g., see Berger & Piore 1980), Japanese dualism has been argued to be distinctive in that, as compared to other industrial countries, the wage differential and mobility gap between the two sectors are much wider (Ernst 1986), the interlocking between large and small firms is tighter, and the subcontracting and subsubcontracting hierarchy more far-reaching (Tatsumi 1971).[3]

Corresponding to this dual industry structure, the Japanese labor market is said to be divided into two segments — one for jobs in large, core firms, the other for jobs in the small or medium, peripheral firms (Broadbridge 1966; Chalmers 1989; Ernst 1986). The concept of a dual labor market consisting of "primary" versus "secondary" segments describes the differential employment conditions that workers face in large versus small firms. Male, regular employees in large firms are primary workers employed in the core firm sector who occupy good jobs and enjoy high wages, good promotion prospects, and permanent, stable employment; by contrast, secondary workers, such as temporary employees, part-time workers, older women, or men coming from second-rate schools, are employed in bad jobs in small firms in the periphery, and receive low wages, limited chances of promotion, and minimal job security.

FIRM INTERNAL LABOR MARKETS

Firm internal labor markets (FILMs) are a useful concept for describing why members of core firms and primary labor markets can experience upward mobility and career advancement. FILMs are groups of jobs within organizations that are characterized by: "(1) a job ladder, with (2) entry only at the bottom, and (3) movement up this ladder, which is associated with a progressive development of knowledge or skill" (Althauser & Kalleberg 1981:130). A firm with a structured FILM has a well-developed set of administrative rules and customs that govern promotions, wages, and benefits (Doeringer & Piore 1971). FILMs ensure a stable employment relationship that facilitates the implementation of effective on-the-job training for firm-specific skills necessary for advanced technological production. By contrast, firms that lack FILMs provide little on-the-job training, low wages and benefits, unstable work, and little or no chance for career advancement. Most jobs within unstructured labor markets require few firm-specific skills and can be performed by any worker who holds some general skills.

Some researchers argue that the permanent employment and seniority-based reward systems in Japan should not be seen simply as manifestations of industrial dualism or employer paternalism, but as elements of FILMs (Cole 1972, 1973; Koike 1983). The permanent employment policy was adopted in the 1920s by large firms as a pragmatic means to stabilize their regular work force (Minami 1986) by reducing employee turnover, and lowering the cost of retraining skilled and technical staff (Cole 1979). While the seniority-based reward system is a mechanism to defer wage payment by underpaying younger

workers, it also facilitates on-the-job-training and job rotation by detaching remunerative rewards from one's immediate task. The historical account for the emergence of FILMs in Japan also echoes the significance of firm-specific skill training as a crucial determining factor.

The institutionalization of FILMs coincided with the urgent needs for firm-specific skill training in large corporations during the 1920s. Spurred by the World War I (1914-18) economic boom, big enterprises flourished in heavy, chemical, and machinery manufacturing industries. To remain competitive, they imported new, foreign technologies from many countries (Helvoort 1979). Machine tools and equipment adapted to local factory conditions varied greatly in design, and so were the particular skills required for their operations. With little public vocational training, and the pressing needs for ready-to-use, firm-specific skills, large companies had to establish internal programs for technical training (Levine & Kawada 1980). To arrange for firm-specific skill training and to secure returns from investments in human capital, these firms progressively tightened their control over recruitment, job allocation, and rewards of skilled workers. The traditional master-apprentice relations were co-opted and brought under their bureaucratic managerial control. The once highly mobile master craftsmen, who governed the supply, training, and compensations of skilled labor, received permanent tenure and supervisor titles in the firms, and their apprentices were turned into regular wage employees (Levine 1965). They were then responsible for training new graduates hired directly from schools. Permanent employment and seniority-based reward systems were applied to skilled workers to reduce their mobility across firms and to facilitate training within the firm.

How do Japanese FILMs differ from those, say, in the U.S.? Some claim that a permanent employment guarantee carries more normative weight in Japan. Regular (male) employees expect lifelong employment, particularly from large employers who are presumed to be capable of fulfilling that obligation. Consistent with this claim, we see that large Japanese firms are under considerable social pressure not to lay off employees except as a last resort during recession[4] (Ernst 1986; Rohlen 1979). Compared to the U.S., wage promotion in Japanese FILMs is based on a ranking hierarchy that is more a function of seniority than job content (Sakamoto & Chen 1992). Accordingly, we see that job descriptions and classification are less detailed in Japan (Cole 1979) and job rotation is less rigid (Hashimoto 1990).

OCCUPATIONAL DIMENSION OF LABOR MARKETS

Internal labor markets are not only structured by firms, but also by occupations. (See Althauser & Kalleberg's 1981 discussion of OILMs, Doeringer & Piore's 1971 discussion of craft submarkets and mobility clusters within firms, and Osterman's 1984 distinctions among the "industrial," "craft," and "secondary" subsystems within firms.) "Occupation" refers to technical work activities that are transferable among employers and to skills that are transportable from firm to firm. Hence, work activities that are structured on the occupational principle tend to be bought and sold in labor markets that are external to a particular firm. Both firm internal labor markets (FILMs) and occupational internal labor

markets (OILMs) have the common structural features of an internal labor market: people who hold jobs in internal labor markets can move upwards on job ladders and acquire progressively greater work skills. A difference between them is that job movements in FILMs are usually within a single firm, while those in OILMs are within a single occupation but may often span firms. Another basic difference between FILMs and OILMs is who controls the employment relationship: while FILMs are basically governed by the employers, OILMs are greatly influenced by the incumbents of specific occupations. People with occupational skills that require special talents or long periods of learning often form associations (e.g., professional association, craft union) to protect their shared interests and influence the labor market processes. Some occupational groups directly set a monopolistic price for their service, others affect the market by controlling the supply of their skills. They govern the qualification procedures, specify standards and ranks, and thereby, occupational internal labor markets are developed.

FILMs and OILMs may overlap, since, in some situations, employment relations may involve occupations and firms simultaneously (Bridges & Villemez 1991; DiPrete & McManus 1993). For instance, firms might compete for the loyalty of someone who is in an OILM (such as a lawyer or other professional) by offering opportunities for promotion and advancement within the firm. If someone is concurrently in a FILM and an OILM, however, it is the occupation, rather than the firm, that circumscribes his labor market since he can change firms while continuing to move up an occupationally defined job ladder. By contrast, if someone who is only in a FILM leaves the firm, this in effect means he must enter the new firm's FILM at the bottom, as most ports of entry are at the bottom of job ladders. For this reason, whether or not one is in an OILM will affect his propensity of making moves across firms or occupations.

One feature claimed to be uniquely Japanese is workers' identification with firms rather than occupations. Occupation has not been given much attention as a factor in determining workers' careers and mobility chances, and it is indeed often ignored in analyses of Japanese labor market structures.[5] Low occupation consciousness among the Japanese has been explained by the so-called "groupism" orientation. In the past, Japanese group consciousness was represented by the concept of *ie* — the traditional household institution. In modern days, the *ie* notion has been carried over to the firm — which is conceived as a united, family-like group (Kaneko 1980). As in a family, employees are permanent members of the firm. The relationship between employee and employer is diffused and lifelong. Total commitment and obligation on both sides is expected, and remuneration is awarded according to an employee's length of membership (seniority) and loyalty. Association within the group is exclusive, and identification with outsiders is greatly discouraged. Japanese tend to form groups according to their common *frame* (e.g., a firm), not their *attributes* (e.g., occupational characteristic) (Nakane 1970). The vertical relation (e.g., super- vs. subordinate) is the primary acting principle for groupings and identification, and thereby undercuts the horizontal relation that binds people of similar attributes together (e.g., class formation or occupational connection).[6]

The occupational principle has less importance as a work-structuring form in Japan, especially in its large-firm sector, as is indicated by the low occupational consciousness and vague job classification schemes. Historically, the craft guild tradition was weak as the country was suddenly awakened to industrialization by the jolt of Western intrusion. When large industrial firms began to mushroom in the late 1910s and 1920s, many craftsmen were co-opted into their bureaucratic management. Traditional skilled labor was eventually replaced by internally trained youngsters recruited from schools, and the master-apprentice relations that organized skilled workers along craft lines within large firms were dissolved. Other craftsmen survived by establishing their own businesses or working for small firms in the competitive sector (Jacoby 1979). Craft unions — occupational associations of allied workers — flourished in the upsurge of labor movements during the 1920s and 1930s. However, they were later driven out by the *Sampo* (the Industrial Patriotic Labor Association) movement (1939-45), a patriotic movement orchestrated by the military government to elicit workers' cooperation with management and to support the war through production (Taira 1970). Radical labor movements immediately after the war were again suppressed. Superseding trade unions were enterprise unions, which built on the structure of *Sampo*, uniting all *regular* employees within the *same* company regardless of their occupations (Helvoort 1979). Excluding other temporary employees, workers organized themselves on an enterprise basis, with permanent employment and seniority wage as their main objectives. Unlike before, union memberships were no longer concentrated among small-firm workers, but among large-firm workers. This principle of unionism — an enterprise-based union that unifies labor and management interests by tying employees' benefits to the ultimate survival (thus profit) of their own company — inevitably perpetuates the dual labor market segmentation and limits workers from joining with their counterparts (by occupation or class) across firms.

Despite all this, however, evidence still points to the importance of occupational differences for mobility in Japan. Using the 1975 Social Stratification and Mobility Survey, Cheng (1991) found that the odds of Japanese men's remaining with their first employer depend on firm size *and* occupation; and given that employer changes occur, then mobility distances among employees reflect the socioeconomic status of their occupational types, with rank order strikingly similar to that in the U.S. (Cheng 1995). Moreover, though to a different degree, large and small firms in Japan rely on the external labor market for their recruitment. Regardless of firm size, the types of experienced workers most sought after are production engineers, computer specialists, personnel and general managers, managerial planners and researchers, finance specialists, and sales promoters (Beck & Beck 1994). Midcareer recruitment targets professionals and managerial workers with high-level technical or special skills. While the trend for increased mobility among all workers is apparent,[7] the rise among professionals is remarkable. In the 1980s, across-firm mobility among male engineers, professionals, managers, and official employees increased by more than 10% annually, exceeding the traditional highest mobility group of skilled workers in the production, transportation, and communication industry (Beck & Beck 1994). As indicated in various case studies, Japanese employees, like many of their counterparts in the U.S., are involved assiduously

in attempts to shape their own careers (see Plath 1983). Professionals, because of the empowerment of their specialized knowledge, demonstrate that under some circumstances, the occupational principle of work control can take over from the administrative principle, even in Japan's large-firm sector.

Hypotheses

The distinction between firm and occupational internal labor markets is useful for understanding the dynamics of job mobility and careers. This conceptualization focuses our attention on "episodes" of labor force experiences as a worker progresses through a career, instead of assuming that individuals are static and always belong to a particular labor market. We use the concepts of FILM and OILM to examine the structural elements of Japanese employment practices and how they might be similar to work arrangements in other industrial social settings. We assume that the observed patterns of job mobility are manifestations of the latent labor market structures (DiPrete & McManus 1993) and postulate four specific types of mobility patterns that are generated by the underlying firm and occupational labor market structures (see also Kalleberg & Hudis 1979). Firm size and occupational types are the key variables that designate the locales of these internal labor markets, whereas the labor market structures are empirically manifested by the tendency of their incumbents to make these various types of job shifts. By specifying hypotheses about the observed mobility patterns associated with the concrete firm size and occupational types, we investigate the structure of the Japanese labor market from the firm and occupation internal labor market perspectives. Table 1 summarizes our hypotheses.

FIRM SIZE

Larger firms are more likely than smaller firms to establish and maintain FILMs because larger firms have more stable sales, higher profits, higher productivity, and more technologically sophisticated production. Large firms are thus better able to provide high employment tenure, high wages and benefits, and internal promotion opportunities. Recruitment preferences go to new graduates or young workers, training is mainly on-the-job, and higher positions are filled from within based primarily on seniority. By contrast, small firms are less likely to develop FILMs because of their unstable sales, less capital-intensive production, limited resources, and higher bankruptcy rates. Consequently, small firms are less able to guarantee stable and lifelong employment or to provide many internal promotion opportunities. With less technologically sophisticated production, firm-specific on-the-job training also becomes less essential. As workers can be more readily substituted, a stable employment relationship is not of great concern to the employer, at least not when the labor supply is abundant. These observations suggest that,

Hypothesis 1: Men in larger firms are more likely than those in tiny firms to be in FILMs. Thus, firm size should be positively related to intrafirm mobility and negatively related to interfirm mobility.

TABLE 1: Hypotheses

	Within-Firm Within-Occ. (FILM)	Within-Firm Across-Occ. (FILM)	Across-Firm Within-Occ. (OILM)	Across-Firm Across-Occ. (SLAM)
Firm size	+	+	-	-
Occupations (vs. unskilled workers)				
Professional	+	-	+	-
Managers	+	-	+	-
Clerks	+	+	?	-
Skilled workers	+	-	+	-
Semiskilled workers	?	+	+	-
Unskilled workers	lowest	lowest	lowest	highest

Hypothesis 2: Men in tiny firms are more likely to be in unstructured, secondary labor markets (SLAMs) than men in large firms. Consequently, men employed in tiny firms are more likely to change occupations and firms simultaneously than those in large firms.

PROFESSIONAL AND TECHNICAL OCCUPATIONS

Professional and technical occupations are exemplars of OILMs, since they require high levels of technical skills and professional training, much of which has to be acquired from outside the firm. Firms most often recruit for these occupations from the external labor market, since internal transfers face technical entry barriers. Moreover, these occupations' job skills are relatively highly "portable" and transferable to other firms. Therefore, professional and technical workers tend to remain in the same occupation when they change firms. At the same time, many incumbents of these OILMs are in FILMs as well, particularly in Japan. This suggests our third hypothesis:

Hypothesis 3: Professional and technical workers are more likely to be in OILMs. However, in some situations, these OILMs might overlap with FILMs. Thus, we expect them to be more likely to have across-firm, within-occupation and within-firm, within-occupation transitions than members of occupations who are not in OILMs, such as unskilled workers.

MANAGERIAL AND ADMINISTRATIVE OCCUPATIONS

Members of these occupations tend to have skills that are more firm-specific (idiosyncratic knowledge of the given firm) than those in professional and technical occupations. In addition, promotion tends more often to be from within the firm: probably from lower-level clerical positions (see Osterman's 1984 findings). Promotions are also more likely to be upwards along authority hierarchies, rather than to involve changes in the nature of one's work. Members of these occupations are more likely to be in FILMs, but changing occupations is less likely. This suggests the next hypothesis:

Hypothesis 4: Managerial and administrative employees are more likely to be in FILMs. Some, especially those at higher levels, may also be in OILMs, and be sought after by other firms. Therefore, overall we expect managers and administrators to experience more within-firm, within-occupation transitions, and to some extent, also within-occupation across-firm shifts, than unskilled workers.

CLERICAL WORKERS

Members of these occupations should have general, transferable skills, and low levels of technical skills. However, Japanese men classified as clerks are likely to perform lower-level managerial work, and so many are at the lowest rung of a job ladder that leads to management. Hence, we expect that at least a portion of male clerks are actually in FILMs — with relatively high probability of within-firm transitions, mostly across occupations — and will later on move into managerial jobs. Therefore, we hypothesize that:

Hypothesis 5: Male clerks tend to be in FILMs. Thus, they have higher probability of within-firm, across-occupation transitions than men in unskilled occupations.

SKILLED WORKERS

These occupations have technical skills that are likely to be transferable from one firm to another. In this sense, they comprise OILMs — much like professional and technical workers. We suggest that,

Hypothesis 6: Skilled occupations are likely to be in OILMs. Therefore, skilled workers — more than unskilled — tend to experience across-firm, within-occupation, and to a certain extent, within-firm, within-occupation transitions also.

SEMISKILLED WORKERS

Members of these occupations are somewhat more likely to be in OILMs than unskilled workers. However, their extent of within-firm or within-occupation mobility should be much less than that of professional technical or skilled workers. Hence, we expect that:

Hypothesis 7: Semi-skilled workers are less likely to have across-firm, across-occupation shifts, but higher likelihood of experiencing across-firm, within-occupation transitions than unskilled workers. However, compared with skilled workers or professionals, they have lower probability of within-firm, within-occupation shifts.

UNSKILLED WORKERS

Since these workers are unlikely to be in either kind of ILM and to have few general or firm-specific skills, we hypothesize that they are likely to be in secondary labor markets (SLAMs):

Hypothesis 8: Unskilled workers are less likely to have within-firm shifts, but higher risks of across-firm and across-occupation shifts.

Data and Methods

THE 1975 JAPANESE SOCIAL STRATIFICATION AND MOBILITY SURVEY (SSM)

The work history data reported in the 1975 SSM survey (Tominaga 1975) were used to test these hypotheses about Japan's labor market structures and mobility patterns. Although the data set is nearly two decades old, it is the latest SSM survey available to non-Japanese scholars.[8] Moreover, since much of the popular understandings of the Japanese employment system focused on the 1960s and 1970s, these data may provide a better test of those hypotheses than more recent ones. The 1975 SSM was a representative national survey for a random sample of men aged 20 to 69. It collected complete work histories and information on many job characteristics for 2,724 cases, of which 2,639 had some work history. Even though data were restricted to men, which limits our assessment of gender differences in mobility patterns, since most of the discussions regarding permanent employment and internal labor markets have focused on men, they are useful for assessing hypotheses associated with these employment relations.

In the survey, 27% of the respondents[9] were still in their first job, and 39% were still working for their first employer[10] when interviewed. The numbers of work history spells reported by each respondent ranges from 1 to 15, and 85% of the respondents reported four or fewer work history spells. About 17% of the sample experienced some kind of interruption in their career, such as unemployment, military service, or school. When job spells are distinguished from nonjob spells,[11] we find that the mean number of jobs ever held by all respondents is 2.6. The distribution of actual number of jobs reported[12] shows that 27% of the respondents have had one job, 30% have had two, 20% have had three, and 11% have had four.

Our analyses are restricted to employees' labor force experiences in the nonagricultural sector, since the farm sector and self- or family-employment are less relevant to the concept of firm and occupation labor markets. From the work histories of the 2,639 cases, we found that 2,169 have been in the nonagricultural employee sector. Although some of them were in the agricultural sector, being self- or family-employed at one time or another, we retained their nonagricultural, employee job histories for analysis.[13] Among the 2,169 cases, a total of 3,017 actual job shifts (transition from job to job) are reported during their stay in the nonagricultural employee sector.

MULTINOMIAL LOGISTIC REGRESSION MODELS

We infer the operation of particular labor market mechanisms from an examination of patterns of actual job movements. We postulate four specific types of mobility patterns as generated by the underlying firm and occupational labor market structures. Workers who change jobs could experience one of four mutually exclusive types of transitions: (1) within-firm, within-occupation, (2) within-firm, across-occupation, (3) across-firm, within-occupation, and (4) across-firm, across-occupation.[14] Hence, our dependent variable is an unordered, polychotomous response variable that indicates the occurrence of a *particular type* of job change. We use multinomial logistic regression techniques[15]

to model the relative odds of a specific type of shift as compared to another type.

To specify the model, let m be the total number of possible outcomes in the discrete-categorical response variable; $j = 1, \ldots, m-1$, indexing the occurrence of a particular type of job shift (e.g., $j = 1$ means a within-firm, within-occupation job shift; $j = 2$ means a within-firm, across-occupation job shift, and so on); and m be the base category for comparison (say, an across-firm, across-occupation job shift). Let P_j be the probability of a specific type of job shift. It can be expressed as a function of some explanatory variables X_i, with β'_j as the effect estimates:

$$P_j = \frac{e^{\beta'_j X_i}}{1 + \sum_{j=1}^{m-1} e^{\beta'_j X_i}}$$

and the base category:

$$P_m = \frac{1}{1 + \sum_{j=1}^{m-1} e^{\beta'_j X_i}}$$

Hence, the relative odds of a specific job shift type (versus the base type) is:

$$\frac{P_j}{P_m} = e^{\beta'_j X_i}$$

In its logit form:

$$\ln\left(\frac{P_j}{P_m}\right) = \beta'_j X_i$$

The β'_j estimates can be interpreted as the effects of the explanatory variables on the log odds of having a particular type of job transition relative to the base type.

Variables

MEASURING FILM AND OILM STRUCTURES

Firm size and occupational types are the variables that designate the locales of these firm and occupational labor markets. We assume that FILMs are more prevalent in larger firms, and that OILMs are more likely to be in professional/technical occupations, and among skilled and semiskilled workers.

Firm Size

The firm size classification has six categories: (1) *tiny* firms with fewer than 31 workers; (2) *small* firms with 31-300 workers; (3) *medium* firms with 301-500 workers; (4) *large* firms with 501-1,000 workers; and (5) *very large* firms with more than 1,000 workers. We created dummy variables to represent each of these discrete categories.

Occupational Type

The occupational type variable was based on the detailed three-digit Japanese census occupation codes. We grouped detailed occupations into the following nine types: (1) *professional and technical;* (2) *administrative and managerial;* (3) *clerical;* (4) *sales;* (5) *service;* (6) *agricultural;* (7) *skilled;* (8) *semiskilled,* and (9) *unskilled production.*[16] Since our analysis focuses on job mobility of employees in the nonagricultural sector, we exclude transitions that originate from agricultural occupations and use only eight occupational categories. We created dummy variables to represent each of these occupational categories.

CONTROL VARIABLES

Educational Attainment

Education is measured in number of years of formal schooling. In general, we expect workers with higher education to have more resources and assets with which to bargain for better jobs, within and across firms. Educational attainment and credentials commonly set the upper bound on promotions and the pace of promotion schedule, and allocate people to different career tracks. Net of other factors, workers with more education will generally reach higher positions through internal promotions and will follow a faster promotion schedule. Hence, *ceteris paribus,* we expect educational attainment to have a positive effect on the probabilities of within-firm shifts, within and across occupations.

With regard to across-firm transitions, studies in the U.S. have found that educational attainment is positively related to the propensity for interfirm mobility and thus to shorter job and firm tenures (Sandefur 1981). A worker's education is often treated as an indicator of his human capital, or specifically, of his general skills. However, the positive educational effect on within- and across-firm transitions might be confounded with the effects of being a member of a particular occupation, since occupational membership depends on education. So, we need to control education when evaluating occupational effects on men's type of job shift. A positive educational effect (holding occupation constant) would mean that workers in similar occupations who have more educational training tend to have higher relative odds of experiencing a certain type of job shift than another.

Labor Force Experience at the Time of Job Transition

In Japan, seniority-based rewards and age constraints on external recruitment are generally assumed to be common, so we expect a strong negative effect of labor force experience on the across-firm shifts, particularly across occupations, but a positive effect on within-firm job transitions, especially within occupation.

Period When Job Shifts Occurred

We need to control for period effects in our analysis since the time periods during which the job shifts reported in the SSM survey occurred represent periods when the Japanese economy and society underwent drastic changes. We divided the time period into: (1) before 1946 (pre-World War II); (2) 1946-55

(Allied Occupation, economic reconstruction and recovery); (3) 1956-65 (economic expansion); and (4) 1966-75 (further modernization). Considering the rapid economic advancement, acute labor demands, and increased implementation of permanent employment in the late 1960s and early 1970s, we expect an overall trend toward more within-firm movements during the postwar era.

Industry

We need to consider industrial differences in evaluating the effects of occupational and firm labor market structures. Industries, based on the 2-digit census codes, are grouped into seven sectors: (1) mining, (2) manufacturing, (3) finance, (4) trade, (5) communication, (6) service, and (7) government.[17] We control for industry in examining job shift patterns along the occupational and firm dimensions.

Results

Among the 2,169 cases who were employed in the nonagricultural sector, 48% experienced some across-firm, across-occupation job shifts; 22% across-firm but within-occupation shifts; 14% within-firm, within-occupation shifts; and 12% within-firm but across-occupation shifts. Table 2 shows average individual characteristics associated with each type of job shift at the time of transition.

On average, across-firm transitions occur at a younger age than within-firm transitions (about age 28 compared to age 33). This is also reflected in the difference of average labor force experiences at time of transition — about 10 years for across-firm shifts compared to 13 years for within-firm shifts. The discrepancy in terms of tenure between within-firm and across-firm transitions is even greater. Employees who experience within-firm transitions, either within or across occupation, have on average about 10 to 11 years of seniority. For those who move across firms, seniority in the previous firm tends to be much less, around 5.5 years. The majority of within-firm transitions are promotions, particularly those that are also within-occupation (90%),[18] compared to only about a quarter (24% to 28%) of across-firm transitions. Not surprisingly, about half (51%) of across-firm, within-occupation transitions, and 43% of across-firm, across-occupation moves originated from tiny firms with fewer than 31 workers, as compared to about 18% of within-firm type of job shifts. Similarly, while about 30% of both types of within-firm transitions occurred in huge firms with more than 1,000 workers, only 10% of across-firm, within-occupation shifts and 16% of across-firm, across-occupation shifts occurred there.

Workers who make within-firm job moves have a higher mean education than those who make across-firm job moves (12 compared to 10 years of education). Occupations associated with across-firm, across-occupation transitions also, on average, have much less prestige than those associated with within-firm, within-occupation moves. Compared across types, within-firm, within-occupation transitions tend to involve a higher percentage of professionals and technicians than others. On the other hand, across-firm, within-occupation moves have a higher percentage of skilled workers. Within-firm shifts,

TABLE 2: Average Individual Characteristics Associated with Each Type of Job Shift at Time of Transition[a]

	N (Shifts)	Age	Labor Force Experience	Tenure Duration	Education (Year)	Occ. Prestige[b]
Within-firm, within-occupation	361	33.0 (.428)	13.5 (.426)	10.21 (.354)	2.2 (.155)	48.2 (.495)
Within-firm, across-occupation	311	32.6 (.504)	13.4 (.480)	11.2 (.452)	11.8 (.164)	45.1 (.570)
Across-firm, within-occupation	713	28.0 (.333)	10.4 (.300)	5.6 (2.10)	10.5 (.099)	43.4 (.348)
Across-firm, across-occupation	1,632	27.5 (.241)	9.8 (.223)	5.7 (.148)	10.3 (.065)	40.8 (.231)

regardless of whether they are across- or within-occupations, tend to involve clerical types of occupations.

Table 3 presents the results of multinomial logit regression analyses of our hypotheses, and reports contrasts between pairs of mobility types. Coefficient estimates show the effects of the explanatory variables on the relative (log) odds of experiencing a certain type of job transition compared to another, if a worker changes his job.[19] The results for the control variables are consistent with our expectations: education and labor force experience have positive effects on the relative odds of experiencing within-firm shifts (both within- or across-occupation) as opposed to both types of across-firm shifts, and an increasingly higher relative odds of within-firm movements (vs. across-firm transitions) during 1956-1965 and 1966-75 as compared to the prewar era. The only industry that differs involves mining, which facilitates across-firm *within-occupation* (as opposed to across-firm *across-occupation*) transitions, and government, which encourages within-firm, within- or across occupational shifts rather than across-firm within-occupational changes. Overall, the occupational and firm mobility patterns are stable with or without controlling for the period and industry effects.

HYPOTHESIS 1

Consistent with our first hypothesis, Table 3 indicates that men in larger firms are more likely than those in tiny firms (with less than 31 workers) to experience within-firm, as opposed to across-firm, mobility. The firm size effects are quite stable as suggested by the positive coefficients associated with all the within-firm shift versus across-firm shift contrasts (columns 1, 2, 4, and 5). For instance, among those who change jobs, the odds of a worker in a huge firm (1,000+) experiencing a *within-firm, within-occupation* job shift as opposed to an *across-firm, across-occupation* shift (column 1) are 3.0 times higher than those of

TABLE 2: Average Individual Characteristics Associated with Each Type of Job Shift at Time of Transition[a] (Continued)

	Percentage Originated in					Percent Promotion
	Professional	Clerk	Skilled	Tiny Firm	Huge Firm	
Within-firm, within-occupation	12.7 (1.8)	38.0 (2.6)	18.3 (2.0)	17.7 (2.0)	30.7 (2.5)	90.3 (1.6)
Within-firm, across-occupation	4.5 (1.2)	42.4 (2.8)	13.8 (2.0)	18.0 (2.2)	29.1 (2.6)	58.8 (2.8)
Across-firm, within-occupation	6.6 (.9)	16.3 (1.4)	32.1 (1.7)	50.8 (1.9)	10.2 (1.2)	24.2 (1.6)
Across-firm, across-occupation	5.0 (.5)	13.0 (.8)	27.1 (1.1)	42.6 (1.3)	16.0 (0.9)	27.9 (1.1)

[a] Standard errors are in parentheses.
[b] Naoi's occupational prestige scores range from 23.4 to 87.3. (For coding scheme, see Tominaga, Naoi & Imada 1983.)

a tiny-firm worker (exp[1.096]). When contrasted to the probability of an *across-firm*, within-occupation shift, his odds of having a *within-firm*, within-occupation shift instead (column 4) are 6.2 times higher than that of a tiny-firm worker (exp[1.814]). We see similar patterns for the *within-firm*, across-occupation versus *across-firm*, across-occupation contrast (column 2), and the *within-firm, across-occupation* versus *across-firm, within-occupation* contrast (column 5). The relative odds of the first contrast, comparing huge-firm workers to tiny-firm workers, are 2.7 (exp[.976]). As for the second contrast, the ratio is 5.4 (exp[1.694]). Such a tendency of mobility is also evident when comparing big firm (501-1,000) and medium firm (301-500) workers with tiny-firm workers. Even workers in small firms (31-300) have higher odds of having within-firm, as opposed to between-firm, job transitions when compared to tiny-firm workers, although to a lesser extent. These findings support the hypothesis that persons in larger firms are more likely than those in tiny firms to be in FILMs.

HYPOTHESIS 2

The evidence just discussed also provides support for our second hypothesis: men in tiny firms are more likely to change occupations and firms simultaneously than men in larger firms. To obtain relative (log) odds of having across-firm, across-occupation shifts instead of within-firm, within-occupation shifts for workers of different firms, we simply change the sign of the coefficients in column 1.[20] For example, the odds of huge-firm workers having across-firm, across-occupation transitions compared to within-firm, within-occupation shifts

TABLE 3: Multinomial Logit Models for Types of Transition, as Employees Change Jobs[a]

	Within-Firm Within-Occ. vs. Across-Firm Across-Occ. (1)	Within-Firm Across-Occ. vs. Across-Firm Across-Occ. (2)	Across-Firm Within-Occ. vs. Across-Firm Across-Occ. (3)	Within-Firm Within-Occ. vs. Across-Firm Within-Occ. (4)	Within-Firm Across-Occ. vs. Across-Firm Within-Occ. (5)	Within-Firm Within-Occ. vs. Within-Firm Across-Occ. (6)
Independent /control variables						
Constant	-5.376***	-4.990***	-1.882***	-3.494***	-3.108***	-.386
	(.757)	(.759)	(.483)	(.826)	(.831)	(.975)
Firm size (vs. less than 31 workers)						
31-300	.586*	.621*	-.180	.765**	.801**	-.035
	(.275)	(.286)	(.174)	(.291)	(.304)	(.362)
301-500	.946*	1.163**	-.048	.994*	1.211**	-.217
	(.408)	(.408)	(.308)	(.435)	(.437)	(.489)
501-1,000	.719	1.039*	-.801+	1.520**	1.840**	-.320
	(.477)	(.463)	(.447)	(.577)	(.568)	(.563)
1,000+	1.096***	.976**	-.718**	1.814***	1.694***	.120
	(.294)	(.308)	(.237)	(.333)	(.348)	(.377)
Occupational type (vs. unskilled workers)						
Prof/tech.	1.547**	-.422	1.634***	-.087	-2.056**	1.970**
	(.557)	(.574)	(.424)	(.627)	(.644)	(.716)
Managers	1.276*	.258	1.678***	-.402	-1.420*	1.018
	(.619)	(.562)	(.495)	(.685)	(.635)	(.711)
Clerks	1.753***	.946*	1.476***	.277	-.531	.807
	(.491)	(.383)	(.348)	(.555)	(.464)	(.563)
Salesmen	.434	.082	.249	.185	-.167	.352
	(.574)	(.462)	(.400)	(.656)	(.561)	(.673)
Skilled	.795	-.306	1.167***	-.373	-1.473**	1.101+
	(.484)	(.393)	(.309)	(.542)	(.463)	(.579)
Semiskilled	.035	-.610	1.214***	-1.179*	-1.824***	.645
	(.542)	(.428)	(.327)	(.599)	(.499)	(.646)
Service	1.248*	-.577	1.258**	-.010	-1.835*	1.825*
	(.582)	(.664)	(.401)	(.650)	(.725)	(.821)

[a] Standard errors are in parentheses.

TABLE 3: Multinomial Logit Models for Types of Transition, as Employees Change Jobs[a] (Continued)

	Within-Firm Within-Occ. vs. Across-Firm Across-Occ. (1)	Within-Firm Across-Occ. vs. Across-Firm Across-Occ. (2)	Across-Firm Within-Occ. vs. Across-Firm Across-Occ. (3)	Within-Firm Within-Occ vs. Across-Firm Within-Occ (4)	Within-Firm Across-Occ vs. Across-Firm Within-Occ (5)	Within-Firm Within-Occ. vs. Within-Firm Across-Occ. (6)
Education (year)	.114**	.113**	.023	.091+	.090+	.001
	(.041)	(.043)	(.034)	(.046)	(.048)	(.051)
Labor force exp (year)	.038***	.043***	.010	.027*	.032*	-.005
	(.011)	(.012)	(.009)	(.012)	(.013)	(.013)
Period (vs. pre-1946)						
1946-1955	-.182	.496	-.358	.177	.854	-.679
	(.495)	(.569)	(.266)	(.526)	(.597)	(.710)
1956-1965	.688	1.312*	-.006	.694	1.318*	-.622
	(.439)	(.525)	(.243)	(.464)	(.549)	(.640)
1966-1975	1.000*	1.083*	-.117	1.116*	1.200*	-.080
	(.429)	(.527)	(.249)	(.457)	(.551)	(.635)
Industry (vs. trade)						
Mining	.021	-.080	.604*	-.584	-.684	.100
	(.422)	(.437)	(.274)	(.440)	(.459)	(.537)
Manufacturing	-.066	-.101	-.294	.227	.192	.035
	(.311)	(.322)	(.239)	(.341)	(.354)	(.380)
Finance	.434	.450	-.369	.803	.819	-.016
	(.523)	(.518)	(.628)	(.673)	(.674)	(.546)
Comm.	-.052	.387	.169	-.221	.218	-.439
	(.418)	(.395)	(.305)	(.452)	(.434)	(.485)
Service	.116	-.223	-.091	.208	-.132	.339
	(.385)	(.470)	(.292)	(.413)	(.496)	(.529)
Government	.527	.495	-.913**	1.440***	1.407**	.032
	(.381)	(.411)	(.317)	(.431)	(.461)	(.488)

+ $p < 10$ * $p < .05$ ** $p < .01$ *** $p < .001$

are only 33% of those of tiny-firm workers (exp[-1.096] = .334). In reverse, that also means, among those who changed jobs, a tiny-firm worker's odds of having an *across-firm, across-occupation* job shift as opposed to a *within-firm, within-occupation* shift are 3.0 times higher than those of someone originating from a huge firm (exp[1.096]). A similar figure (2.7 = exp[.976], column 2) is obtained for contrast made with *within-firm* but across-occupation shifts. Results in columns 4 and 5 also suggest a consistently higher odds of having *across-firm,* within-occupation type of transitions (contrasted with *within-firm,* within- or

across- occupation transitions) for tiny-firm workers. As opposed to within-firm, within-occupation shifts, their odds of being associated with across-firm, within-occupation transition are about two times those of the small-firm workers, three times the medium-firm workers, five times the big-firm, and six times the huge-firm (exp[.765] = 2.15, exp[.994] = 2.70, exp[1.520] = 4.57), and (exp[1.814] = 6.13, respectively). Similar patterns are observed when relative odds comparisons made between *across-firm, within-occupation* transitions against *within-firm, across-occupation* ones (see column 5). These findings are consistent with our hypothesis that tiny-firm workers are much more likely to be in the unstructured, secondary labor markets (SLAMs).

HYPOTHESIS 3

The data also substantiate the hypothesis that professional and technical employees are more likely than other workers to be in occupational internal labor markets (OILMs). Among men who change jobs, those in professional and technical occupations are less likely than those in unskilled work to change occupations regardless of whether the job shifts are within- or across- firms (Table 3, columns 3 and 6). The odds of professional and technical workers experiencing a within-firm, *within-occupation* transition (as opposed to within-firm, *across-occupation* shift) are about 7.2 times those of unskilled workers (exp[1.970], column 6). Similarly, their odds of experiencing across-firm, *within-occupation* shifts (vs. across-firm, *across-occupation*) are 5.1 times those of unskilled workers (exp[1.634], column 3). In contrast to *across-firm, across-occupation* moves, the odds of professionals and technicians having *within-firm, within-occupation* transitions are 4.7 times higher than those of unskilled employees (exp[1.547], column 1). At the same time, they do not have a higher relative odds of experiencing within-firm, *across-occupation* shifts (vs. across-firm, across-occupation) (column 2). Actually, the relative likelihood for them to make *across-firm, within-occupation* shifts (as opposed to *within-firm, across-occupation*) is significantly higher, about eight times those of unskilled workers (exp[2.056] = 7.8, column 5).

HYPOTHESIS 4

When workers change jobs, managerial and administrative employees have higher odds (about four times, exp[1.276] = 3.58) than unskilled workers of experiencing a *within-firm, within-occupation* transition (vs. *across-firm, across-occupation*) (column 1). This finding seems to suggest that managers and administrators are in FILMs. However, they also tend to stay within the same occupation when moving across firms. Their relative odds of having across-firm, *within-occupation* shifts (vs. across-firm, *across-occupation*) (column 3) are five times higher than the unskilled workers, and similarly are their odds of having *across-firm, within-occupation* shifts (vs. *within-firm, across-occupation*) (exp[1.678] = 5.35 and exp[1.420] = 4.14, column 5, respectively).

HYPOTHESIS 5

Compared to unskilled workers, men in clerical occupations have much higher odds of experiencing both types of intrafirm transitions (within- and across-occupation) as opposed to interfirm, across-occupation transitions (columns 1 and 2). Their relative odds are, respectively, six times and three times greater (exp[1.753] = 5.77 and exp[.946] = 2.58). This is somewhat consistent with our hypothesis that they are in FILMs. However, contrary to our expectation that male Japanese clerks tend to change occupations (becoming managers) as they are promoted within the FILMs, their relative odds of making within-firm, *across-occupation* transitions (vs. within-firm, *within-occupation* shifts) are not higher than those of unskilled workers (in fact, the coefficient, though insignificant, has the opposite sign — see column 6). Moreover, when moving across firms, their relative odds of keeping the same occupation (vs. not) are significantly greater than those of the unskilled (four times, exp[1.476] = 4.38, column 3).

HYPOTHESIS 6

The mobility pattern of skilled workers resembles that of professional and technical workers and substantiate our hypothesis that skilled occupations are likely to be in OILMs. Regardless of whether shifts occurred within or across firms, skilled employees are more likely than the unskilled to remain in the same occupation (columns 3 and 6). Their relative odds of experiencing across-firm, *within-occupation* shifts (vs. across-firm, *across-occupation*), and within-firm, *within-occupation* shifts (vs. within-firm, *across-occupation*), are about three times higher when compared to unskilled workers (exp[1.167] = 3.21 and exp[1.101] = 3.01 respectively). Also, their relative likelihoods of having *across-firm, within-occupation* shifts (vs. *within-firm, across occupation*) are four times higher than those of unskilled workers (exp[1.473] = 4.36, column 5).

HYPOTHESIS 7

Compared to unskilled workers, semiskilled workers have higher odds of across-firm, *within-occupation* transitions: 3.4 times as against across-firm, *across-occupation;* and 6.2 times as against *within-firm, across-occupation* (exp[1.214], column 3, and exp[1.824], column 5, respectively). These findings support our seventh hypothesis that semiskilled workers may also be in OILMs. However, unlike professionals and skilled workers, their relative odds of having *within-firm, within-occupation* moves (vs. *across-firm, across-occupation*) are no higher than those of unskilled workers (column 1); nor do they tend to remain in the same occupation if within-firm shifts ever occur (column 6). They do not seem to be simultaneously in FILMs as are professionals, technicians, and skilled workers.

HYPOTHESIS 8

Our findings are consistent with the hypothesis that unskilled workers are likely to be in SLAMs instead of FILMs. Among job changers, unskilled workers are more likely to have across-firm, across-occupation shifts (as opposed to within-

firm, within-occupation) than men in professional, managerial, clerical, semiskilled production, and service occupations (column 1). Specifically, their relative odds of making across-firm and occupation transitions — versus within-firm and occupation — are about three times those of managers and service workers, five times professionals, and six times clerks (exp[1.276] = 3.58, exp[1.248] = 3.48, exp[1.547] = 4.70, exp[1.753] = 5.77, respectively). Also consistent with our eighth hypothesis is the finding that the odds of unskilled workers changing both firms *and* occupations — versus changing firms but remaining in the same occupation — are much higher than all others except salesmen (column 3). Their relative odds are about three to four times that of semiskilled, skilled, service, and clerical workers, and more than five times that of professionals and managers (exp[1.214] = 3.37, exp[1.167] = 3.21, exp[1.258] = 3.52, exp[1.476] = 4.38, exp[1.634] = 5.12, and exp[1.678] = 5.35, respectively). These facts suggest unskilled workers are likely to be in SLAMs instead of OILMs.

Conclusions

Contrary to the impression often given by popular accounts of Japanese employment relations, our results indicate that Japanese men often do change employers as well as occupations during their careers. Using a conceptualization of internal labor markets that distinguishes between firm (FILMs) and occupational (OILMs) internal labor markets, we tested various hypotheses about how organizational and occupational dimensions were related to four patterns of job shifts.

We found considerable support for our hypotheses. Men in larger organizations were more likely than those in smaller firms to experience within-firm mobility, a pattern consistent with our view that FILMs are more apt to be found in larger firms. Men in tiny firms were more likely to change occupations and firms simultaneously than men in huge firms, suggesting a correspondence between tiny firms and unstructured, secondary labor markets (SLAMs). Workers in unskilled occupations were also apt to be in SLAMs. With regard to OILMs, we found that professional and technical workers were more likely to experience within-occupation mobility. Also suggesting the operation of OILMs were some of our results for managerial and administrative employees, skilled occupations, and semiskilled workers. Moreover, professionals and managers were more likely to have within-firm, within-occupation and across-firm, within-occupation shifts than skilled or semiskilled workers, suggesting high-level white collars in the private sector probably benefit from both types of ILMs. These patterns of job movements are consistent across time periods and industries.

Our findings suggest the utility of systematically examining patterns of job mobility in Japan, rather than relying on simplistic assumptions based on stereotypes of permanent employment or a seniority wage system. With deeper involvements in international trades and operations, and constant needs for product diversification and innovations, large Japanese corporations are under growing pressure to externalize their training and rely more on the external labor market for needed specialized personnel. To accommodate the special

needs of experienced recruits, they inevitably have to modify their employment practices. Considering the aging of the work force and the lower commitment of the younger generation to their employers, the trend toward increased job mobility is prominent. Along with interfirm mobility, we expect that occupation will increase in importance as a work structure shaping labor market processes in Japan.

Future research should compare the mobility of men and women and enrich our quantitative approach with concrete case studies of occupational careers. Studies should also compare Japanese mobility patterns to those in the U.S. and other industrial countries. Our results support the need to reexamine hypotheses about cross-national convergence and divergence in mobility patterns. Research along these lines is likely to enhance markedly our understanding of cross-national variations in the structures and processes of labor market segmentation.

Notes

1. Taira argues that lifetime employment applies only to core employees in private firms with over 500 workers; Marsh and Mannari use the number 1,000. Dore suggests that, in addition to these employees, the system also applies to government firms.

2. Within each industrial sector, relationships between large and small firms are often vertical and unequal in power; large firms dominate small ones through "parent-child" relations in the subcontracting system.

3. The tighter and more far-reaching subcontracting and interlocking relationships between large and small firms are partly historical outcomes of government intervention. In the early and mid-1940s, Japan's military government instituted a controlled economy to support the war efforts. Large firms took in small ones as subcontractors, forming a system of enterprise groupings sustaining the munitions industries (Fujita 1971). After the war, subcontracting was revived and reintroduced as *Kigyo Keiretsu*, an integrated series of companies — a management mechanism to use existing capital and human resources efficiently. In 1955, large enterprises began to dominate small firms by interlocking them through the selective alignment and subcontracting processes (Tatsumi 1971).

4. Other personnel reduction strategies include internal transfer, retraining, and reallocation, layoff of temporary workers and termination of subcontracts, delaying new recruitment, and encouraging early retirements with special pay.

5. Tokunaga (1984) in describing the Japanese labor market structure even suggests that there is "no division of blue-collar workers into 'skilled,' 'semiskilled,' and 'unskilled' categories" (27). In American stratification and mobility studies, by contrast, occupation has been recognized as a major work structure with consequential impact on work life and mobility patterns, and its effect is above and beyond that of organization (e.g., Kalleberg 1983; Kalleberg & Berg 1987).

6. While such a notion is perhaps intuitive in understanding informal group formations among the Japanese, it lacks concrete explanations of how and why the primacy of frame and vertical relations have persisted in the process of industrialization. It has not addressed the facts that, within the same firm, there are different career tracks among workers (e.g., white collars vs. blue collars), and that across firms of various sizes, different labor market structures (e.g., internal vs. external) exist. Although organization affects workers' career prospects and mobility tendency in many important ways, we should not take it as the *sole* factor, particularly when commitment to and identification with one's own firm in fact varies according to firm size, work performance, and occupations (Lincoln & Kalleberg 1990; Mouer & Sugimoto 1986).

7. Government statistics show, over the years 1962 to 1987, the percentage of men who were continuously working but seeking new or additional employment had doubled from 4% to 8%.The percentage wanting to change jobs rose from 4% to 10% (Japan Statistics Bureau 1990). Correspondingly, research studies have reported increases in mid-career recruitment, even for

large-firm regular employee positions (e.g., Watanabe 1987). Beck and Beck (1994) have also captured this mobility trend in their presentations of job movements and intentions of some large-firm, prime-age, top executives.

8. The 1985 SSM survey contains not only men's, but also women's, work data. However, the data have not been made available to foreign scholars. For further details on the sampling procedures and data collection of 1975 SSM survey, see Ishida 1993.

9. Respondents include all persons who reported at least one work history spell (all those who have had some kind of labor force experience).

10. The difference between the two percentages indicates the percentage of respondents who are still with their first employer and have experienced some kind of within-firm job shift (they changed jobs but not employers).

11. Nonjob spells are spells during which respondents are either unemployed, in military service, or have gone back to school. Job spells indicate periods of time in which respondents are actually holding jobs, either as self-employed or as employees.

12. Changing jobs does not necessarily mean changing employers (since this could reflect within-firm job shifts). However, changing employers implies changing jobs. Therefore, men who have had two or more jobs might still work for the same employer.

13. The nonagricultural, employee restriction made here only applies to the state of job origin but not the job shift destination. In other words, the selection of job shifts for analysis is contingent only upon the origin, not the destination. It is conceivable that some of the across-firm shifts accompany a change in employment status, and some across-firm across-occupation shifts involve movements into the agricultural sector.

14. Within- versus across-firm mobility indicates whether or not the job shift is associated with a change in employer. Within- versus across-occupation movements are gauged by using the 288, detailed, three-digit Japanese census occupation codes.

15. For an introduction of multinomial logit models, see Maddala 1983. We use the MLOGIT procedure (SAS version 5.3) to estimate these models.

16. We elaborate the eight-category occupational classification used by the SSM researchers by differentiating "service workers" (identified by the three-digit, 288 occupational titles in the Japanese census classification) from other major categories. The service worker category includes protective service workers (e.g., policemen, firefighters, and guards), domestic, and nondomestic service workers (e.g., cooks, waiters, and garbage collectors). Professionals are researchers, engineers, physicians, lawyers, teachers, artists, etc. Managerial and administrative workers are officials, directors, and managers in public and private enterprises. Clerks engage in accounting, billing, typing, mailing, and other communication. Salesmen are dealers (retail or wholesale), brokers, and sales agents. Agricultural workers include farmers, livestock raisers, foresters, loggers, gardeners, and fishermen. Skilled workers are involved in product manufacturing such as assemblers and repairmen for machines, automobiles, other transportation equipment, precision and optical instruments, and traditional craftsmen making kimonos, tatami, pottery, and traditional food products. Semiskilled workers are machine operators, metal processing workers, electric machine/appliance assemblers and repairmen, and manufacturing workers in textile, paper, wood, rubber, and cement products. Unskilled workers are mostly laborers, diggers, and carriers. The complete occupational classification scheme is available on request from the first author.

17. Two sources in the survey tell whether the respondent was in the government at job change: (1) he was reported as working in a government firm when asked about his employer's firm size (and no further size information is given for a certain government agency), or (2) he was recorded as being in local or national government in the industry question. In either case, we consider this person as being in the government industrial category, and whoever is reported as government industry will not be in other firm size categories except government firm type. The coefficients associated with the dummy government variable can be interpreted as contrasts made to the trade workers (the base category for industry), or to the tiny-firm workers (the base category for firm size).

18. The SSM survey did not ask directly whether each job shift was a promotion. However, it provides some information on the managerial ranking of each of the positions. Here promotion is operationally defined by comparing respondents' positions before and after the job shift. Shifts from ordinary employee rank to any position with a managerial title are considered promotions. Also, managerial titles are hierarchically ranked: (1) department, station, administration directors; (2) section managers, auditors, advisors; (3) branch, store, or office managers; (4) head clerks, chiefs, assistant section managers; and (5) foremen, supervisors. Moving from a lower to a higher level constitutes a promotion, as does shifting from the rank of "foremen/supervisors" to "school principals, deans, directors" or "chief researchers/technicians/editors" in educational, health, or research institutes.

19. In analysis not presented here, we examined this set of hypotheses using a slightly different methodological framework, though we reach strikingly similar conclusions. Using a discrete-time event history approach, we estimated multinomial logistic regression models of the competing risks of the four mutually exclusive types of transitions (within-firm within-occupation, within-firm across-occupation, across-firm within-occupation, and across-firm across-occupation). We allowed each job shift type to have its own hazard function dependence on time (i.e., employment duration) and a similar set of explanatory and control variables. When evaluating the propensities of moving (vs. not moving), we can summarize our findings as follows. Employment duration is positively associated with within-firm transition but negatively with across-firm transition rates. Education is positively related to all shifts except the across-firm across-occupation type. Medium to large firms tend to provide more within-firm opportunities, while huge enterprises and government sector tend to discourage across-firm movements. Compared to unskilled workers, professionals, managers, and skilled workers are much less likely to make across-occupation shifts, within or across firms. While clerks are unlikely to make across-firm across-occupation shifts, they are much more prone to within-firm within-occupation shift. When contrasting the competing risks of making different types of job shifts (e.g., within-firm within-occupation vs. across-firm across-occupation, within-firm across-occupation vs. across-firm across-occupation, etc.), the coefficients associated with the firm size and occupational type variables are very much similar to the results presented here in Table 3.

20. Any pairs of contrast can be made from results shown in Table 3 simply by taking the difference between the coefficients, as long as the original sets of contrasts share the same reference category. For example, coefficients in column 6 can be obtained, within rounding errors, from those in column 1 minus the corresponding estimates in column 2.

References

Abegglen, James C. 1958. *The Japanese Factory: Aspects of Its Social Organization.* Free Press.

Althauser, Robert P., and Arne L. Kalleberg. 1981. "Firms, Occupations, and the Structure of Labor Markets: A Conceptual Analysis." Pp. 119-49 in *Sociological Perspectives on Labor Markets,* edited by Ivar Berg. Academic Press.

Aoki, Masahiko. 1984. "Aspects of the Japanese Firm." Pp. 3-43 in *The Economic Analysis of the Japanese Firm,* edited by Masahiko Aoki. North Holland.

Beck, John C., and Martha N. Beck. 1994. *The Change of a Lifetime: Employment Patterns among Japan's Managerial Elite.* University of Hawaii Press.

Berger, Suzanne, and Michael J. Piore. 1980. *Dualism and Discontinuity in Industrial Societies.* Cambridge University Press.

Bridges, William P., and Wayne J. Villemez. 1991. "Employment Relations and the Labor Market: Integrating Institutional and Market Perspectives." *American Sociological Review* 56:748-64.

Brinton, Mary C., Hang-Yue Ngo, and Kumiko Shibuya. 1991. "Gendered Mobility Patterns in Industrial Economics: The Case of Japan." *Social Science Quarterly* 72:807-16.

Broadbridge, Seymour A. 1966. *Industrial Dualism in Japan: A Problem of Economic Growth and Structural Change.* Aldine.

Chalmers, Norma J. 1989. *Industrial Relations in Japan: The Peripheral Workforce.* Routledge; Nissan Institute of Japanese Studies, University of Oxford.

Cheng, [Mariah] Man Tsun. 1991. "The Japanese Permanent Employment System: Empirical Findings." *Work and Occupations* 18:148-71.

___. 1995. "Employment Transitions in the Japanese Male Labor Force: A Log-Multiplicative Analysis of Mobility Structures." *Work and Occupations* 22:188-214.

Cole, Robert E. 1972. "Permanent Employment in Japan: Facts and Fantasies." *Industrial and Labor Relations Review* 26:615-30.

___. 1973. "Functional Alternatives and Economic Development: An Empirical Example of Permanent Employment in Japan." *American Sociological Review* 38:424-38.

___. 1979. *Work, Mobility, and Participation: A Comparative Study of American and Japanese Industry.* University of California Press.

DiPrete, Thomas A. and Patricia A. McManus. 1993. "Tenure, Mobility and Incumbency: Comparing Observed Patterns of Earnings with Predictions from an Elaborated Theory of Occupational and Firm Labor Markets." *Research in Social Stratification and Mobility* 12:45-82.

Doeringer, Peter B., and Michael J. Piore. 1971. *Internal Labor Markets and Manpower Analysis.* Heath.

Dore, Ronald P. 1973. *British Factory, Japanese Factory: The Origins of National Diversity in Industrial Relations.* University of California Press.

Ernst, Angelika. 1986. "Employment, Wages, and Export Competitiveness." Pp. 209-29 in *Japan's Response to Crisis and Change in the World Economy,* edited by Michèle Schmiegelow. M.E. Sharpe.

Fujita, Keizo. 1971. "Short History of Small Enterprises in Japan." Pp. 10-19 in *Small Business in Japan's Economic Progress,* edited by Tokutaro Yamanaka. Asahi Evening News.

Hashimoto, Masanori. 1990. "Employment and Wage Systems in Japan and Their Implications for Productivity." Pp. 245-94 in *Paying for Productivity: A Look at the Evidence,* edited by Alan S. Blinder. The Brookings Institution.

Hashimoto, Masanori, and John Raisian. 1985. "Employment Tenure and Earnings Profiles in Japan and the United States." *American Economic Review* 75:721-35.

Helvoort, Ernest van. 1979. *The Japanese Working Man: What Choice? What Reward?* University of British Columbia Press.

Ishida, Hiroshi. 1993. *Social Mobility in Contemporary Japan: Educational Credentials, Class and the Labor Market in A Cross-National Perspective.* Stanford University Press.

Jacoby, Sanford. 1979. "The Origins of Internal Labor Markets in Japan." *Industrial Relations* 18:184-96.

Japan Statistics Bureau. 1990. *Japan Statistical Yearbook 1990.* Management and Coordination Agency.

Kalleberg, Arne L. 1983. "Work and Stratification: Structural Perspectives." *Work and Occupations* 10:251-59.

Kalleberg, Arne L., and Ivar Berg. 1987. *Work and Industry: Structures, Markets, and Processes.* Plenum Press.

Kalleberg, Arne L., and Paula M. Hudis. 1979. "Wage Change in the Late Career: A Model for the Outcomes of Job Sequences." *Social Science Research* 8:16-40.

Kaneko, Yoshio. 1980. "The Future of the Fixed-Age Retirement System." Pp. 104-23 in *The Labor Market in Japan,* edited by Nishikawa Shunsaku and translated by Ross Mouer. University of Tokyo Press.

Koike, Kazuo. 1983. "Internal Labor Markets: Workers in Large Firms." Pp. 29-61 in *Contemporary Industrial Relations in Japan*, edited by Taishiro Shirai. University of Wisconsin Press.

Levine, Solomon B. 1965. "Labor Markets and Collective Bargaining in Japan." Pp. 633-67 in *The State and Economic Enterprise in Japan: Essays in the Political Economy of Growth*, edited by William W. Lockwood. Princeton University Press.

____. 1983. "Careers and Mobility in Japan's Labor markets." Pp.18-33 in *Work and Lifecourse in Japan*, edited by David W. Plath. SUNY Press.

Levine, Solomon B., and Hisashi Kawada. 1980. *Human Resources in Japanese Industrial Development*. Princeton University Press.

Lincoln, James R., and Arne L. Kalleberg. 1990. *Culture, Control, and Commitment: A Study of Work Organization and Work Attitudes in the United States and Japan*. Cambridge University Press.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.

Marsh, Robert V., and Hiroshi Mannari. 1971. "Lifetime Commitment in Japan: Roles, Norms, and Values." *American Journal of Sociology* 76:795-812.

Minami, Ryoshin. 1986. *The Economic Development of Japan: A Quantitative Study*, translated by Ralph Thompson and Ryoshin Minami. Studies in the Modern Japanese Economy. St. Martin's Press.

Mincer, Jacob, and Yoshio Higuchi. 1988. "Wage Structures and Labor Turnover in the United States and Japan." *Journal of Japanese and International Economies* 2:97-133.

Mouer, Ross E. and Yoshio Sugimoto. 1986. *Images of Japanese Society: A Study in the Social Construction of Reality*. Japanese Studies. KPI.

Nakamura, Takafusa. 1981. *The Postwar Japanese Economy: Its Development and Structure*, translated by Jacqueline Kaminski. University of Tokyo Press.

Nakane, Chie. 1970. *Japanese Society*. University of California Press.

Odaka, Kunio. 1986. *Japanese Management: A Forward-Looking Analysis*. Asian Productivity Organization.

Osterman, Paul. 1984. "White-Collar Internal Labor Markets." Pp. 163-89 in *Internal Labor Markets*, edited by Paul Osterman. MIT Press.

Plath, David W. (ed.). 1983 *Work and Lifecourse in Japan*. SUNY Press.

Rohlen, Thomas P. 1979. "'Permanent Employment' Faces Recession, Slow Growth, and an Aging Work Force." *Journal of Japanese Studies* 5:235-72.

Sakamoto, Arthur, and Meichu D. Chen. 1992. "Earnings Inequality and Labor Market Segmentation in Japan and the United States." Paper presented at the 1992 American Sociological Association Meetings, Pittsburgh.

Sandefur, Gary D. 1981. "Organizational Boundaries and Upward Job Shifts." *Social Science Research* 10:67-82.

Shirai, Taishiro (ed.). 1983. *Contemporary Industrial Relations in Japan*. University of Wisconsin Press.

Tachibanaki, Toshiaki. 1984. "Labor Mobility and Job Tenure." Pp. 77-102 in *The Economic Analysis of The Japanese Firm*, edited by Mashiko Aoki. North Holland.

Taira, Koji. 1962. "Characteristics of Japanese Labor Markets." *Economic Development and Cultural Change* 10:150-68.

___. 1970. *Economic Development and the Labor Market in Japan*. Columbia University Press.

Tatsumi, Nobuharu. 1971. "Small-Medium Enterprises and Big Corporations." Pp. 74-83 in *Small Business in Japan's Economic Progress*, edited by Tokutaro Yamanaka. Asahi Evening News.

Tokunaga, Shigeyoshi. 1984. "The Structure of the Japanese Labor Market." Pp. 25-55 in *Industrial Relations in Transition: The Case of Japan and the Federal Republic of Germany*, edited by Shigeyoshi Tokunaga and Joachim Bergmann. University of Tokyo Press.

Tominaga, Ken'ichi. 1975. *Social Stratification and Mobility Survey, 1975* [Machine-Readable Data File], Institute for Social Science Research, University of California, Los Angeles (Distributor).

Tominaga, Ken'ichi, Atsushi Naoi, and Takatoshi Imada. 1980. "Current Trends in Studies of Social Stratification and Social Mobility in Japan." Paper presented at the Japan-U.S. Conference on Social Stratification and Mobility, Hawaii.

Vogel, Ezra F. 1979. *Japan as Number One: Lessons for America*. Harvard University Press.

Watanabe, Shin. 1987. *Job-Searching: A Comparative Study of Male Employment Relations in the United States and Japan*. Ph.D. diss., University of California, Los Angeles.

Dynamics of Organizational Responses to Competition*

JOEL A.C. BAUM, *University of Toronto*
JITENDRA V. SINGH, *University of Pennsylvania*

Abstract

This research examines how processes of adaptation and selection operate jointly in the evolution of a population of day-care centers (DCCs) in metropolitan Toronto. We study how DCCs alter their organizational niches, defined by productive capacities and targeted resources, in response to competition and how these changes influence their survival chances. Exhibiting little structural inertia, DCCs modified their organizational niches in response to changing competitive conditions, often without any harmful effects. Indeed, DCCs that moved to less competitive organizational niches improved their survival chances. At the same time, however, competition increased rates of DCC failure and the organizational niche changes made by DCCs did not, on average, affect either the intensity of competition they faced or their longevity. We discuss the implications of these findings, which indicate that the evolution of the DCC population is a joint function of adaptation and selection processes.

The nature and consequences of change among existing organizations and the manner in which populations of organizations and industries evolve over time underlie a current polemic for researchers in organization theory. Although adaptation theories vary in the extent to which they rely on the rational calculation and the capabilities of organizational actors, most suggest that relatively pliable organizations commonly respond to threats and opportunities created by environmental change by altering organizational strategies and structures in ways that enhance performance and prospects for survival (e.g., Chandler 1977; Pfeffer & Salancik 1978; Porter 1980; Rumelt 1986; Thompson 1967). Consequently, for adaptationists, variability among organiza-

** This research was supported in part by the Office of Research Administration, University of Toronto, and the Sol C. Snider Entrepreneurial Center, the Wharton School, University of Pennsylvania. For helpful comments on an earlier version of this article we are grateful to Howard Aldrich, Jennifer Bethel, Ned Bowman, Filip Caeldries, Jacques Delacroix, Kathy Hick, Helaine Korn, Bill Starbuck, and Anand Swaminathan. We are also grateful to Debbie Freeman, Brian Gaon, Gayle Greenbaum, and Sonja Saksida for assistance with data collection and coding.* *Direct correspondence to Joel A.C. Baum, Div. of Mgmt. and Econ. University of Toronto, Scarborough Campus, 1265 Military Trail, Scarborough, Ontario M1C 1A4.*

 Social Forces, June 1996, 74(4):1261-1297

tions over time reflects the cumulative changes of individual organizations. In contrast, selection theories depict organizations as relatively inert entities for which adaptive response is not only difficult and infrequent, but hazardous as well (Amburgey, Kelly & Barnett 1993; Hannan & Freeman 1977, 1984). Consequently, for selectionists, organizational variability is primarily a reflection of relatively inert organizations replacing each other.

Although adaptation and selection theories are often portrayed as mutually exclusive alternatives with distinct implications (Scott 1992), they are sometimes seen as complementary (Bourgeois 1984; Child 1972; Hrebiniak & Joyce 1985; Levinthal 1991). Building upon their potential complementarity, researchers have begun to examine organizational flexibility empirically and to model the adaptiveness of organizational change dynamically (for a review, see Baum 1996).

The purpose of this study of organizational niche changes (i.e., moves by ongoing organizations within a resource space) in a population of day-care centers (DCCs) in metropolitan Toronto is to extend research specifying how processes of adaptation and selection jointly influence organizational evolution in two ways. First, we examine the influence of organizational niche overlap and nonoverlap on rates at which DCCs change organizational niches. Second, we study the adaptiveness of organizational niche changes by examining whether DCCs that move to organizational niches with less intense competition have different fates than those that change to organizational niches where competition is more intense. Despite the centrality of competition to adaptationist and selectionist accounts of organizational change, research on the dynamics of organizational change has not yet examined how the consequences of organizational change are related to their effects on the intensity of competition.

This study is part of a research program in which we investigate the implications of organizational niches for understanding population dynamics. In addition to the "macro-niches" of organizational populations (Hannan & Freeman 1977, 1989), organizations have their own "micro-niches" defined by the intersection of resource requirements and productive capacities at the organizational level (McKelvey 1982:109-10). Thus, the macro-niches of organizational populations encompass multiple organizational niches. In our research, we define the organizational niches of DCCs on the basis of a critical resource dimension: the ages of children DCCs are licensed to enroll. We measure organizational niche change as the expansion or contraction of the range of ages DCCs are licensed to enroll. Organizational niche changes, which are moves by organizations within the macro-niche, entail organization-level changes in organizational capabilities and assets.

In earlier studies (Baum & Singh 1994a, 1994b), we found that, because DCCs operating in different organizational niches face different competitive landscapes, different patterns and rates of founding accompany different choices of organizational niche at founding and that, after founding, DCCs occupying different organizational niches endure different fates. The theoretical arguments in this article broaden the scope of our work to include organizational niche change and its survival consequences.

For multiple reasons, DCC organizational niche change provides a useful mechanism by which to investigate organizational responses to competition and

their survival consequences. Previous studies of DCC founding and failure established that competition plays a significant role in the dynamics of this population (Baum & Singh 1994a, 1994b). Organizational niche change is the main way DCCs change their core features — the claims they stake for themselves in terms of the clients they target, the services they have the capacity to provide, the complexity and differentiation of internal organization, the human and material resources they employ, and their linkages with clients and the external environment. Therefore, organizational niche change represents a potentially disruptive form of change that selectionist theory suggests will subject DCCs to a significant risk of failure in the short run (Hannan & Freeman 1984). The typical smallness of DCCs reinforces further their high risk of death from change (Hannan & Freeman 1984). On the other hand, adaptationist theories view organizational niche changes as pervasive, planned, and adaptive responses to competition that influence organizational success (Fligstein & Dauber 1989; Miles & Snow 1978; Porter 1980; Ramanujam & Varadarajan 1989; Rumelt 1986). Indeed, given the organizational simplicity of DCCs and their environments, one might expect them to possess a high potential for developing adaptive responses to competition.

Organizational Responses to Competition

In organizational ecology, the potential for competition among organizations is predicted mostly to be a function of similarity in resource requirements: the more similar the resource requirements, the greater the potential for intense competition (Hannan & Freeman 1977, 1989; McPherson 1983). At one extreme, organizations with identical resource requirements are perfect competitors. At the other, organizations with resource requirements that do not intersect at all do not compete. In general, the potential for competition between any two organizations is directly proportional to the overlap or intersection of their organizational niches. More formally, the potential for competition between two organizations in two organizational niches i and j can be represented by an *organizational niche overlap weight*, w_{ij}, that varies between 0 (i.e., no overlap in organizational niches and no potential for competition) and 1 (i.e., complete overlap in organizational niches and a high potential for competition) (McPherson 1983). The organizational niche overlap weight, w_{ij}, which gives the potential competition experienced by an organization in organizational niche i resulting from the presence of an organization in organizational niche j, is equal to the proportion of the resources of organizations in organizational niche i that are also required by those in organizational niche j. This ecological approach gives a quite simple way of measuring the potential competition for the underlying resources between different members of an organizational population or industry.

OVERLAP DENSITY

For every organization, we define a variable, *overlap density,* which captures the potential competition faced by that organization. Overlap density measures the potential for competition by aggregating the organizational niche overlaps of a focal organization with those of all other members of its population. More formally, the overlap density for an organization in organizational niche i at time t is given by

$$\textit{Overlap Density}_{it} = N_{it} + \sum_{j \neq i} w_{ij} N_{jt}, \tag{1}$$

where N_{it} is the number of population members in organizational niche i at time t, N_{jt} is the number of organizations in organizational niche j at time t, and w_{ij} is the organizational niche overlap weight for organizations in organizational niche i with those in organizational niche j. Earlier research shows that increases in overlap density have competitive effects that lower rates of founding and increase rates of failure (Baum & Singh 1994a, 1994b). In this article we examine the effects of overlap density on rates of organizational niche change.

How do organizations respond to competition? Swaminathan and Delacroix (1991) suggest that organizations do not normally respond to competition by *dying* (Hannan & Freeman 1977, 1989) but by engaging in search for alternative ways to sustain themselves. Kelly and Amburgey's (1991) study of the U.S. airline industry led them to conclude similarly that managers may possess more discretion regarding change than selectionists typically propose. Following Hawley (1950), Swaminathan and Delacroix argue that organizations become differentiated as competition pushes less fit competitors out of the market, and losing organizations are transformed either through territorial or functional differentiation by entrepreneurs seeking out distinct functions in which they hold a competitive advantage. Recently, this view has been supported in studies of the California wine industry (Delacroix, Swaminathan & Solt 1989; Swaminathan & Delacroix 1991), providing evidence that differentiation occurred in response to competition and that lateral migration into a neighboring market niche (table wines) lowered failure rates for California wineries. Singh, Tucker, and Meinhard's (1991) finding that rates of change in a population of voluntary social service organizations responded to changes in both institutional and competitive conditions, and Haveman's (1994) finding that diversification activities of California S&Ls were responsive to changes in the competitive environment further corroborate this position.

Since greater overlap density implies greater interorganizational competition for resources, the adaptationist view implies that overlap density is positively related to the intensity of search by organizations for alternative ways to sustain themselves. Both expansion and contraction of an organization in the niches it occupies may serve as responses to the problem of increased competition: diversification into new market segments can moderate the effects of competition and withdrawal from a market segment can avoid competition. Therefore, we hypothesize:

Hypothesis 1. Overlap density is positively related to the rate of organizational niche expansion and contraction.

NONOVERLAP DENSITY

Whereas overlap density captures the potential for competition between organizations by aggregating resource overlaps, the difference between overlap density and population density (Hannan & Freeman 1989) is the aggregate of the resource nonoverlaps. We refer to this difference as *nonoverlap density* (Baum & Singh 1994a, 1994b). Nonoverlap density is an estimate of the number of organizations whose resource requirements do not overlap with those of the focal organization. Thus, for each organization, overlap density aggregates the organizational niche overlaps with all other organizations in the population and captures the potential for competition, and nonoverlap density aggregates the organizational niche nonoverlaps and captures information about noncompeting organizations. In terms of equation 1

$$\textit{Nonoverlap Density}_{it} = N_t - [N_{it} + \sum_{j \neq i} w_{ij} N_{jt}], \qquad (2)$$

where N_t is the number of organizations in the population (i.e., population density) at time t and all other terms are as defined in equation 1.

For organizations considering organizational niche expansion, nonoverlap density provides information on potential competition in market segments into which they might diversify. Mitchell (1989) showed how such potential competition lowered rates of entry into emerging subfields by incumbent firms in the U.S. diagnostic imaging industry. Haveman (1994) found similar effects of potential competition on rates of entry by California S&Ls into new markets. Since greater nonoverlap density implies greater potential competition for resources in *other* market segments that may be opportunities for expansion, we expect nonoverlap density to be negatively related to rates of organizational niche expansion.

Our expectations about the effect of nonoverlap density on rates of market contraction are based on a quite different logic. Differentiation that separates organizational resource requirements reduces competition (Baum & Mezias 1992; Chamberlin 1933; Delacroix, Swaminathan & Solt 1989; Hannan, Ranger-Moore & Banaszak-Holl 1990). In addition, differentiation can create complementary functional differences between organizations (Barnett 1990; Barnett & Carroll 1987; Fombrun 1986; Hawley 1950). When such differences arise, they can lead to direct mutualistic relations among organizations. For example, DCCs licensed to enroll infant children may create future demand for those licensed to enroll toddlers. Or they can cooperate directly by referring potential customers to each other. Nonoverlap density may also produce diffuse mutualism by capturing information about the widening social acceptance of the organizational form itself (Hannan & Carroll 1992; Hannan & Freeman 1989). Given the lack of competition for the underlying resources and the potential facilitative effects of mutual demand enhancement and the widening social acceptance of the organizational form, we expect nonoverlap density to foster mutualistic relations among organizations. Supporting this expectation, earlier research shows that increases in nonoverlap density increase rates of DCC founding and lower rates of DCC failure (Baum & Singh 1994a, 1994b). We expect that the mutualism produced by increases in nonoverlap density will help sustain organizations in their current organizational niches, reducing the

need for withdrawal from a market segment. Based on these arguments and this evidence, we hypothesize

Hypothesis 2. Nonoverlap density is negatively related to the rate of organizational niche expansion and contraction.

Competition, Organizational Change, and Failure

Research on organizational change in organization and management theory often focuses on the content of changes: A change to a more advantageous configuration is considered as adaptive, while a switch to a less advantageous configuration is considered detrimental (Amburgey, Kelly & Barnett 1993). Complementing this focus on content, Hannan and Freeman's (1984) structural inertia theory offers a model of the process of organizational change that considers both internal and external constraints on change.

STRUCTURAL INERTIA THEORY

Perhaps the most striking aspect of structural inertia theory is the relationship proposed between organizational change and the liability of newness, the propensity of young organizations to have higher failure rates (Stinchcombe 1965). Hannan and Freeman (1984) propose that some changes produce a renewed liability of newness by robbing an organization's history of survival value. Change renews an organization's liability from newness when it reshuffles work groups, reshapes communication lines and role relationships, and requires the use of new or altered operating routines. Such change renders obsolete an organization's accumulated skills, roles, and routines, reducing its reliability of performance back toward that of a new organization. Changes that upset an organization's exchange relationships with the environment or raise questions about its legitimacy can also reset an organization's liability-of-newness-clock and raise its failure rate.

However, Hannan and Freeman (1984) do not expect the disruptive effects of a given change to be the same for all organizations. Organizational age and size are both predicted to affect the likelihood of surviving the short-run shock of a core change attempt. Because their internal structures and routines are more institutionalized and their external linkages are better established, older organizations are especially likely to experience disruption as a result of core change. In contrast, larger organizations, although less likely to attempt core changes in the first place, are less likely to die during a core change attempt. Largeness can buffer organizations from the disruptive effects of core change by helping, for example, to maintain both old and new ways of doing things during the transition or to overcome short-term deprivations and competitive challenges that accompany the change attempt.

From these arguments, Hannan and Freeman (1984) conclude that, frequently, actions to promote survival expose organizations to a short-run increased risk of failure equivalent to a liability of newness. However, if an organization survives the short-run shock of change, the risk of failure is expected to decline with the passage of time as performance reliability is reestablished, external relationships are restabilized, and organizational

legitimacy is reaffirmed. If decline in the death rate subsequent to the change continues at a rate identical to that before the change, the organization will face the short-term risk of change without any long-term benefit. But if the decline in the death rate after the change is slower than before, the organization will increase both its short-term and long-term risk of failure. Only if the decline is faster than the original rate will the organization benefit in the long run from taking on the short-term risks of change. Thus, from a selection perspective, although organizational change is viewed as disruptive in the short run, it can, ultimately, be adaptive in the long run if the organization manages to overcome the problems associated with the initial disruption (Amburgey, Kelly & Barnett 1993).

Supporting structural inertia theory predictions, in a study of the effects of organizational change on rates of failure in the Finnish newspaper industry, Amburgey, Kelly, and Barnett (1993) found detrimental short-run effects on organizational survival for changes in content and publication frequency that declined as the change receded into the past. Amburgey, Kelly, and Barnett also found that the disruptive effects of changes in content and publication frequency increased with the age of newspaper organizations. Lacking information on organizational size, Amburgey, Kelly, and Barnett could not examine size variation in the disruptive effects of change. In a comparable study, of presidential and managerial succession in the early Iowa telephone industry, Haveman (1993) found detrimental short-run effects of leadership succession on organizational survival that declined as the succession receded into the past. However, the disruptive effects of leadership changes did not increase with the age of telephone companies. Although Haveman did have information on organizational size, she did not examine size variation in the disruptive effects of succession. And, finally, in a study of California wineries, Delacroix and Swaminathan (1991) found no evidence of detrimental effects for brand and product changes in either the short or long run, although they did not examine either age or size variation in the disruptive effects of brand and product changes. Given the scarcity, equivocality, and incompleteness of previous tests of Hannan and Freeman's (1984) predictions, we test the following hypotheses about immediate and dynamic effects of organizational niche change on the rate of failure, which capture the structural inertia theory specification more fully:

Hypothesis 3. The failure rate increases immediately after organizational niche expansion or contraction but declines with the passage of time.

Hypothesis 4. The disruptive effects of organizational niche expansion and contraction increase with the age of the organization.

Hypothesis 5. The disruptive effects of organizational niche expansion and contraction decrease with the size of the organization.

EFFECTS OF CHANGE ON COMPETITION

The emphasis of structural inertia theory on change processes has resulted in less attention being given by ecological researchers to the content of changes. Although ecological researchers do differentiate broad categories of changes according to their content (e.g., content versus frequency change, brand versus

product change, presidential versus managerial change), all instances of a particular category of change are assumed to have the same effect on organizational survival. However, there may be within-category differences in change content that alter their survival implications markedly. One such within-category difference is variation in the effects of changes on competition. Do the disruptive effects of organizational niche changes depend on how they influence the intensity of competition that organizations face?

We propose that changes to organizational niches in which competition is more intense will be detrimental to prospects of organizational survival and that changes to organizational niches in which competition is less intense will improve organizational survival chances. Shifting to a more competitive organizational niche may increase the risk of failure for several reasons. Intense competition creates resource scarcity (Carroll & Hannan 1989). When resources are scarce, organizations cannot move quickly to full-scale operation and may find it difficult to minimize performance disruptions by, for example, maintaining both old and new ways of doing things during periods of transition. Intense competition also results in tight niche packing (Carroll & Hannan 1989). When organizational niches are crowded, few resources go unexploited. This forces new entrants, which are at a disadvantage in head-to-head competition with incumbents, to use inferior or marginal resources, making new client relationships more difficult to establish. Organizations that move into more competitive organizational niches may also be more vulnerable to questions about organizational legitimacy resulting from changes in their visible missions. These deprivations and challenges during the initial period after a change into a more competitive organizational niche may be difficult to recover from and, as a result, may affect the viability of the organization throughout its existence (Carroll & Hannan 1989). Thus, moving into a more competitive organizational niche may magnify the disruptive effects of change, while moving into a less competitive organizational niche may moderate any disruption. Since, as we discussed earlier, greater overlap density implies greater competition, we expect moves to organizational niches with higher levels of overlap density to increase failure rates.[1]

Hypothesis 6. The failure rate is positively related to increases in overlap density that result from organizational niche expansion and contraction.

Research Methods

DATA DESCRIPTION

We compiled life history information from two sources on all licensed DCCs in metropolitan Toronto from 1971 to 1989. The Community Information Center of Metropolitan Toronto's *Directory of Day Cares and Nursery Schools,* published yearly since 1971, provided detailed information on all DCCs in metropolitan Toronto. The Ontario Ministry of Community and Social Services' *Day Nurseries Information System,* which, starting in 1979, contains annual licensing records for all DCCs in Ontario, provided a cross reference to the *Directory*. Between 1971 and 1989, 682 DCCs were founded. The founding date was defined as the year in which the original operating license was issued by the ministry or the year

of first appearance in the *Directory*. Of the DCCs founded since 1971, 139 ceased operations by the end of 1989. The failure date was defined as the first year of reported closing in the ministry archives or *Directory*. At the start of 1971, there were 109 DCCs in metropolitan Toronto. Of these left-censored DCCs, 39 failed by the end of 1989. Through telephone interviews and searches of the Ontario Incorporation Indexes and telephone directories, founding dates were verified for all but two left-censored DCCs. Therefore, the analysis included 789 DCCs, of which 176 failed during the observation period.

DCC Organizational Niches

DCCs in metropolitan Toronto are licensed by the Ministry of Community and Social Services to provide full-time, collective care to five or more children in one or more of the following age ranges: *infant*, birth to 18 months; *toddler*, 19-30 months; *preschool*, 31-60 months; and *school-age*, 61-120 months. DCCs are free to become licensed in any or all age ranges but must pass an inspection performed by provincial authorities to ensure that specified licensing criteria are met before enrolling children in a particular age range. In addition to varying in the ages of children they may legally enroll, DCCs serving children of different ages also differ on a variety of organizational dimensions. To obtain a license to serve a given number of children in a particular age range, a DCC must possess a particular configuration of staff, program, equipment, and facility requirements geared to the developmental needs of children in the age range. Therefore, the configurations of human and material resources possessed by DCCs differ according to the age ranges served. DCCs licensed in more than one age range are required to allocate these configurations of resources exclusively to the care of children in the specified age groups. As a result, DCCs licensed in several age ranges possess more diverse and complex configurations of human and material resources. Thus, a DCC's licensed age ranges provides information on its care-giving capabilities as well as its formally targeted client-market domains. Using information on licensed ages ranges, we identified ten DCC organizational niches. Table 1 shows the number of DCCs occupying each organizational niche in each observation year.[2]

DCC Organizational Niche Expansion and Contraction

For DCCs, we define organizational niche expansion as the addition of a new licensed age range and organizational niche contraction as the withdrawal from a previously licensed age range. Organizational niche expansion requires new investments in human and material resources to meet licensing requirements.[3] Organizational niche expansion also requires DCCs to cope with additional burdens of increased diversification including increased structural complexity, greater human and material resource diversity, and a broader and more complex task environment. Organizational niche expansion also requires DCCs to cultivate new market linkages, communicate new organizational goals to key stakeholders, and be cautious in how they allocate slack resources. By comparison, organizational niche contraction lowers the cost of maintaining a DCC, reduces organizational and environmental complexity, and potentially frees resources for alternative uses. Organizational niche contraction may therefore

TABLE 1: Number of DCCs in Each Organizational Niche, 1971-89

	Organizational Niche[a]										
Year	I	T	P	S	IT	TP	PS	ITP	TPS	ITPS	DCC[b]
1971	0	0	47	2	3	0	48	1	3	5	109
1972	0	0	49	2	5	0	50	2	2	7	117
1973	0	0	51	4	7	0	54	3	1	11	131
1974	0	0	60	5	8	1	57	4	0	15	150
1975	1	0	80	6	13	4	61	8	0	11	184
1976	1	0	98	23	15	4	74	9	1	11	236
1977	1	0	107	27	12	2	81	12	1	11	254
1978	1	0	113	27	14	2	79	11	4	10	264
1979	1	1	124	46	11	3	82	15	8	6	297
1980	1	1	131	60	10	6	87	17	9	7	329
1981	1	1	135	67	10	9	91	14	11	8	347
1982	1	1	151	75	12	11	101	16	14	8	390
1983	1	0	152	81	14	18	107	18	15	8	414
1984	0	0	160	86	15	23	111	21	17	7	440
1985	0	0	154	87	15	40	115	29	23	9	472
1986	1	1	153	91	18	54	132	35	28	13	526
1987	2	1	158	97	20	63	134	47	42	15	579
1988	2	2	164	104	23	65	138	49	44	14	605
1989	3	7	158	103	19	68	141	59	39	16	613

[a] Organizational niches:
I = Infant, 0-18 months
T = Toddler, 19-30 months
P = Preschool, 31-60 months
S = School-age, 61-120 months
TPS = 19-120 months
TP = 19-60 months
PS = 31-120 months
ITP = 0-60 months
IT = 0-30 months
ITPS = 0-120 months.

[b] DCC: Total number of organizations

serve as a cost-reduction measure for DCCs unable to maintain wide organizational niches (i.e., more diversified strategies). However, because it disrupts child care for some families, organizational niche contraction is likely to provoke questions from a DCC's constituents about its legitimacy, accountability, and future reliability.

Thus, organizational niche expansion and contraction alter the core features of DCCs (Baum 1990) — the claims they stake for themselves in terms of the clients they target, the services they have the capacity to provide, the complexity and differentiation of internal organization, the human and material resources they employ, and their linkages with the external environment. Notably, organizational niche expansion appears likely to produce both internal and

external disruptive implications for DCCs, while the disruptive effects of organizational niche contraction appear primarily external.

From archival information, the organizational niche expansion and contraction activities of each DCC were noted annually in its historical data record. DCCs underwent 277 organizational niche expansions and 147 organizational niche contractions during the observation period. It is indicative of the significance of organizational niche changes for DCCs that, although these changes occurred with moderate frequency in the population, most DCCs underwent few, if any, changes. In addition, all changes involved either the addition or the deletion of a single age range adjacent to those already served. Table 2 shows the frequency of expansions and contractions by origin and destination organizational niches.

ANALYSIS

We tested the hypotheses using parametric models of organizational change and failure. A multivariate point-process approach (Amburgey 1986) was used to model the competing risks of organizational niche change and failure. In the analysis, organizational niche expansion, contraction, and failure were modeled as competing risks. In other words, in each year of its existence a DCC was treated as having a competing risk of expanding its organizational niche, contracting its organizational niche, or disbanding. Operationally, we specified each marginal process in terms of the instantaneous transition rate:

$$r_{jk}(t) = \lim q_j(t, t+\Delta t)/\Delta t, \tag{3}$$

where q_j is the discrete probability of organization j experiencing event type k (i.e., expansion, contraction, or failure) between t and $t + \Delta t$, conditional on being at risk for the event at t. Rate models have the advantages of being flexible in specification and reducing the effects of the right-censoring problem (Tuma & Hannan 1984). The models we estimated were of the following general form

$$r_k(t) = \exp\ (\beta_k X_t), \tag{4}$$

where X_t is a vector of the values of the time-varying covariates and β_k is a vector of parameter estimates indicating the effects of the covariates on the rate of occurrence. In this model, rates of change and failure are assumed to be loglinear functions of the covariates, which precludes negative predicted rates. We used *RATE* (Tuma 1980) to estimate the vector of parameters β_k by the method of maximum likelihood.

Since DCCs licensed in all four age ranges cannot, by definition, expand their organizational niches, the sample for the analysis of organizational niche expansion includes only DCCs at risk of expanding their organizational niches between t and $t + \Delta t$ (i.e., licensed in three or fewer age ranges). Conversely, since DCCs licensed in only one age range cannot, by definition, contract their organizational niches, the sample for the analysis of organizational niche contraction includes only DCCs at risk of contracting their organizational niches between t and $t + \Delta t$ (i.e., licensed in two or more age ranges).

TABLE 2: DCC Organizational Niche Expansions and Contractions by Organizational Niche[a]

To Org. Niche j[b]	From Organizational Niche i[b] I	T	P	S	IT	TP	PS	ITP	TPS	ITPS
I					(2)					
T					(2)	(5)				
P						(17)	(50)			
S							(9)			
IT	5	4						(3)		
TP		7	40					(5)	(16)	
PS			87	16					(12)	
ITP					9	23				(18)
TPS						17	34			(8)
ITPS								14	21	

[a] Numbers in parentheses are frequencies of organizational niche contractions.
[b] Organizational niches:

I = Infant, 0-18 months	TP = 19-60 months
T = Toddler, 19-30 months	PS = 31-120 months
P = Preschool, 31-60 months	ITP = 0-60 months
S = School-age, 61-120 months	TPS = 19-120 months
IT = 0-30 months	ITPS = 0-120 months.

INDEPENDENT VARIABLES

Overlap Density and Nonoverlap Density

To test hypotheses 1 and 2, we computed values for overlap density and nonoverlap density for each DCC in each year using equations 1 and 2. Values for overlap density were computed by summing yearly values for the number of DCCs in an organizational niche (see Table 1) multiplied by the appropriate organizational niche overlap weights, w_{ij} (see Appendix B). For example, using equation 1, the overlap density (rounded to the nearest number of competitors) for DCCs in the infant-toddler (IT) organizational niche in 1981 is

$$Overlap\ density_{(IT)1981} = (.6 \times 1_I) + (.4 \times 1_T) + (1 \times 10_{IT}) + (.4 \times 9_{TP}) + (1 \times 14_{ITP}) + (.4 \times 11_{TPS}) + (1 \times 8_{ITPS}) = 41$$

where $N_I \ldots N_{ITPS}$ are the number of DCCs in each organizational niche in 1981. The value for nonoverlap density equals the difference between the total number of DCCs and overlap density (347 - 41 = 306). Yearly values for overlap density for all organizational niches are given in Appendix C. Since earlier research has shown that overlap and nonoverlap densities significantly influence failure rates in the DCC population (Baum & Singh 1994a), we also include these variables in the failure analysis.

Local and Diffuse Overlap and Nonoverlap Densities

In the DCC population, clients tend to come from geographically contiguous areas at a higher rate than from distant places. Since day-care services are provided on the premises of DCCs, parents must transport their children to and from DCCs daily. This transportation cost localizes demand, partitioning the market for day care into numerous local geographic segments. As a result, the location of a DCC influences the potential for organizational niche-based competition among DCCs (Baum & Singh 1994a, 1994b). Therefore, we computed separate local and diffuse measures of overlap density and non-overlap density based on the geographic location of a DCC within metropolitan Toronto. Local densities were calculated for each DCC using information on DCCs operating in the same constituent city of metropolitan Toronto as the focal DCC.[4] Conversely, diffuse densities were computed using information on DCCs operating in all other constituent cities.

Organizational Niche Change and Time-Clock Variables

To model the immediate and dynamic effects of organizational niche changes on rates of DCC failure predicted in hypothesis 3, we followed the procedure outlined by Amburgey, Kelly, and Barnett (1993) and used by Delacroix and Swaminathan (1991). First, we created two dummy variables. The first dummy variable was set equal to zero until the occurrence of an organizational niche expansion. Thereafter, the variable was set equal to one. The second dummy variable was constructed in a similar manner from the occurrence of organizational niche contractions. In addition, we created two time-clock variables to track the elapsed time since a DCC last expanded or contracted its organizational niche. The time-clock variables were set equal to zero until an organizational niche change of either type occurred. After an organizational niche expansion, the organizational niche expansion time-clock variable measured the natural logarithm of the number of years since the change occurred. The counter was reset after any subsequent expansions. The organizational niche contraction time-clock variable was derived in an identical manner from the occurrence of organizational niche contractions. In combination, the organizational niche change and time-clock variables capture the immediate and longer-term effects of organizational niche changes on the risk of DCC failure, permitting the short-term and longer-term effects of organizational niche changes to be different.

To test hypothesis 4, following Amburgey, Kelly, and Barnett (1993), we allowed the dummy variables for organizational niche expansion and contraction to interact with organizational age. These interactions equal zero before the first change and equal organizational age thereafter. If the first change occurs early on, age is lower and the change in the failure rate smaller; if the first change occurs later on, age is higher and the failure rate is greater. To test hypothesis 5, we allowed the dummy variables for organizational niche expansion and contraction to interact with organizational size at the time of the change. These interactions equal zero before the first change and equal organizational size at the time of the change thereafter.

Overlap Density Change

To test hypothesis 6, we calculated the change in the intensity of competition resulting from each organizational niche change. We did this by comparing values of overlap density for a DCC's new and old organizational niches at the start of the year in which it changed. The effects of organizational niche changes on overlap density are a joint function of the resulting changes in organizational niche width and in the resource overlaps with potential competitors. For example, expanding from the preschool (P) organizational niche into the preschool-school-age (PS) organizational niche in 1985 would change overlap density as follows:

$$\textit{Overlap density}_{(P)1985} = (1 \times 154_P) + (1 \times 40_{TP}) + (1 \times 115_{PS}) + (1 \times 29_{ITP}) + (1 \times 23_{TPS}) + (1 \times 9_{ITPS}) = 370$$

$$\textit{Overlap density}_{(PS)1985} = (.33 \times 154_P) + (.67 \times 87_S) + (.33 \times 40_{TP}) + (1 \times 115_{PS}) + (.33 \times 29_{ITP}) + (1 \times 23_{TPS}) + (1 \times 9_{ITPS}) = 279$$

where $N_I \ldots N_{ITPS}$ are the number of DCCs in each organizational niche in 1985 (see Table 1). Overall, the move from preschool to preschool-school-age adds 87 potential school-age (S) competitors but, by increasing the organizational niche width from 30 months to 90 months (see Appendix A), lowers the resource overlaps with DCCs in all organizational niches other than those that include both the preschool and school-age ranges.

Using information on the effects of organizational niche changes on overlap density, we created two variables. The first variable, *Δ overlap density (expansion),* was set equal to zero until the occurrence of an organizational niche expansion. Thereafter, the variable was set equal to the change in overlap density resulting from the expansion. After any subsequent expansion, the variable was reset to the value of the new change in overlap density. The second variable, *Δ overlap density (contraction),* was constructed in a similar manner from the changes in overlap density resulting from organizational niche contractions. We also computed separate variables to capture the effects of organizational niche changes on local and diffuse overlap densities.[5]

CONTROL VARIABLES

To rule out plausible alternative explanations we also controlled for the effects of several organizational characteristics and the environmental carrying capacity.

Organizational Characteristics

Organizational age was defined as the natural logarithm of the number of years since the year of founding. *Organizational size* was measured for each DCC in each year as the natural logarithm of the licensed capacity specified in its annual operating license, which sets the maximum number of children that a DCC can enroll legally. We also included changes in the licensed capacity of each DCC (lagged by one year) in the analysis. Because these changes may reflect discrepancies between demand for services and organizational capacity, this variable may also capture some effects of recent performance on rates of change and failure (Scott 1992). The *profit orientation* of DCCs was also controlled using a dummy variable coded 1 for nonprofit DCCs and zero for profit-

oriented DCCs because influential institutional actors and the public in general have increasingly opposed for-profit child care in metropolitan Toronto and elsewhere in Canada (Maynard 1985).

We also controlled for two kinds of institutional linkages, *purchase-of-service agreements* (POSAs), which link DCCs to the Metropolitan Toronto Children's Services Division (CSD), and *site-sharing arrangements* (SSAs), which involves operating in space leased from or provided by community-based organizations including community centers, schools, and churches. These linkages can influence rates of organizational change and failure by conferring legitimacy and more specific resources on organizations (Baum & Oliver 1991; Miner, Amburgey & Stearns 1990). The effects of POSAs and SSAs were controlled with two dummy variables coded one for DCCs with POSAs or SSAs and zero otherwise.

To control for the possibility that organizational experience with a particular type of change increases the probability of repeating the change but that the probability declines as the change recedes into the past (Amburgey, Kelly & Barnett 1993; Delacroix & Swaminathan 1991; Ginsberg & Baum 1994), we created two variables to count the cumulative occurrence of organizational niche expansion and contraction. These variables were set equal to zero until the first change, 1 until the second change, 2 until the third change, and so on. To control for the time-lapse effect, we included the organizational niche change time-clock variables described above in the change models.

We also control for variation in rates of change and failure by organizational niche using four age-range dummy variables. For each year of a DCC's existence, the dummy variables corresponding to the age ranges it was licensed to serve were coded 1. All others were coded zero. Finally, to examine if left-censored DCCs had systematically different rates of change and failure, a dummy variable coded 1 for DCCs founded before 1971 and zero otherwise was included in the analysis.

Environmental Carrying Capacity

We estimated the demand for DCCs licensed to enroll children in age range i in location j on a yearly basis using the following equation:

$$Demand_{ij} = EN_{ij} + \%WW(NC_{ij} - EN_{ij}) + \%SP[(NC_{ij} - EN_{ij}) \times (1 - \%WW)] \quad (5)$$

where $Demand_{ij}$ is the estimated annual demand for DCC services in age range i in location j, EN_{ij} is the number of children in age range i in location j that are enrolled in a DCC, $\%WW$ is the percentage of women residing in metropolitan Toronto that participate in the work force full-time, NC_{ij} is the total number of children in age range i in location j, and $\%SP$ is the percentage of children from birth to 10 years of age in metropolitan Toronto that reside in single-parent households.[6] In the equation, demand comprises three additive components: the number of children enrolled in DCCs; the number of children with working mothers not enrolled in DCCs; and the number of children living in single-parent households not enrolled in DCCs. We weight the final component by ($1 - \%WW$) since the vast majority of single parents in metropolitan Toronto are working women (City of Toronto Planning and Development Department 1983; Social Planning Council of Metropolitan Toronto 1981). Thus, the demand equation weights the number

of children by factors related to the likelihood that they will require day care. The annual estimates for each age range in each location were used to construct estimates for total and local organizational niche demand for each DCC based on the age ranges it was licensed to enroll at the start of each year and its geographic location. To control for a possible *demand-pull* effect on rates of organizational niche change, we also included a variable measuring the demand for services in the age range(s) each DCC was not licensed to serve at the beginning of each year.

In addition, we controlled for (1) the interest rate for small business loans, (2) the value of the Children's Services Division budget in constant dollars, which measures the government resources provided to parents in need for purchasing services from DCCs with POSAs, (3) the number of early childhood education graduates from Ontario community colleges, which represent the pool of new labor for DCCs that meets the staff qualification requirements, and (4) the number of foundings and failures in a DCC's organizational niche in the prior year in the failure models (Delacroix, Swaminathan & Solt 1989).

Results

ORGANIZATIONAL NICHE CHANGE

Table 3 reports the maximum-likelihood estimates for the analysis of rates of DCC organizational niche expansion and contraction. Models 1 and 2 estimate the effects of overlap density, nonoverlap density, and the organizational and environmental control variables on the rate of organizational niche expansion and contraction, respectively. Supporting hypotheses 1 and 2, the rate of organizational niche expansion increased significantly with increases in overlap density and declined with increases in nonoverlap density. Thus, DCCs were *pushed* into new market segments by competition in their current organizational niche and *pulled* into new market segments by the prospect of low potential competition in the new market segment. The estimates for organizational niche contraction in model 2 provide only partial support. Contrary to hypothesis 1, the rate of organizational niche contraction is negatively related to overlap density. However, hypothesis 2 is supported by the significant negative estimate for the effect of nonoverlap density on the rate of contraction. Thus, DCCs responded to increasing competition by expanding their organizational niches but not by contracting them. One possible explanation for this finding is that in a competitive environment, contraction hampers a DCC's competitive position relative to other DCCs, since most parents want to keep their children in the same DCC and not switch when their children reach a new age range. On the other hand, in the absence of competition, a DCC can contract without fear of losing enrollment in its remaining age ranges to DCCs that cover a broader range of ages.

The estimates for local and diffuse disaggregations of overlap density and nonoverlap density in models 3 and 4 are broadly parallel. However, only diffuse nonoverlap density fails to affect the rate of market contraction significantly. Thus the rate of organizational niche contraction was lowered by the presence of DCCs in the same city but in different organizational niches. This suggests the operation of a direct form of mutualism based on complemen-

tary functional differences among DCCs in different organizational niches, for example, infant DCCs generating future demand for toddler DCCs, or DCCs in different organizational niches referring potential customers directly to each other. Separate local and diffuse effects of overlap density suggest that competition among DCCs entailed both geographically localized competition for similar clients and diffuse competition for client-specific resources that are not location-specific. One such resource is financial assistance provided by the Children's Services Division to city residents who qualify for financial assistance, which is budgeted and allocated on the basis of the four age ranges defined in the Day Nurseries Act. Another client-specific resource is skilled labor that is specialized in care for children of particular ages. However, notably, the coefficients for local overlap density and local nonoverlap density are larger than those for their diffuse counterparts. This means that neighboring DCCs generally had larger influences on each other's fates than DCCs operating in different cities. This is consistent with our expectation that interactions among DCCs would be stronger at the local level.

Several of the control variables also influenced rates of DCC organizational niche change significantly. In contrast to earlier studies, the probability of future organizational niche expansion is lowered further by the occurrence of successive organizational niche expansions but increases with the amount of time since the last organizational niche expansion. The probability of future organizational niche contraction is also lowered further by the occurrence of successive organizational niche contractions, but unrelated to the amount of time since the last contraction. By comparison, the likelihood of future expansion and contraction both increase with successive organizational niche changes of the opposing type and decline with the amount of time since the last opposing change. One possible explanation for these effects is the limited number of opportunities available to DCCs for organizational niche expansion and contraction — that is, each organizational niche expansion reduces the options for further expansion and increases options for contractrions, and each contraction has the corresponding opposite effect.

The coefficients for age and size provide little support for structural inertia arguments. Although rates of organizational niche expansion decrease significantly with size, they increase significantly with age, consistent with Singh, Tucker, and Meinhard's (1988) *fluidity of aging* thesis. The rate of organizational niche contraction is unrelated to organizational age but is significantly higher for larger DCCs. Growing DCCs have significantly higher rates of organizational niche expansion and significantly lower rates of organizational niche contraction. DCCs with site-sharing arrangements (SSAs) were more likely to expand their organizational niches and significantly less likely to contract their organizational niches. Nonprofit DCCs were significantly less likely to either expand or contract their organizational niches, possibly because profit fails to motivate nonprofit DCCs to expand into lucrative market segments or withdraw from unprofitable ones.

The coefficients for the age range dummy variables indicate that adding toddler and preschool age ranges increased the rate of further expansion but adding infants lowered the rate. In contrast, dropping any age range increased the future probability of contraction. Notably, these age range effects generally

TABLE 3: Maximum Likelihood Models of DCC Organizational Niche Change[a]

	1 Expansion[b]	2 Contraction[c]	3 Expansion[b]	4 Contraction[c]
Independent variables				
Overlap density	.021*	-.009*		
	(.004)	(.003)		
Nonoverlap density	-.020*	-.006*		
	(.005)	(.003)		
Diffuse overlap density			.019*	-.008*
			(.004)	(.003)
Local overlap density			.028*	-.023*
			(.007)	(.007)
Diffuse nonoverlap density			-.018*	-.003
			(.006)	(.002)
Local nonoverlap density			-.030*	-.008*
			(.009)	(.003)
Cumulative org. niche expansion	-3.663*	2.192*	-3.634*	2.190*
	(.622)	(.415)	(.619)	(.417)
Org. niche expansion time-clock	1.113*	-.448*	1.103*	-.444*
	(.315)	(.252)	(.313)	(.253)
Cumulative org. niche contraction	1.338*	-2.356*	1.354*	-2.351*
	(.385)	(.693)	(.388)	(.696)
Org. niche contraction time-clock	-.364*	.491	-.371*	.499
	(.209)	(.357)	(.211)	(.358)
ln(organizational age)	.995*	.213	.989*	.207
	(.136)	(.149)	(.135)	(.149)
ln(organizational size)	-.517*	1.604*	-.538*	1.604*
	(.175)	(.220)	(.175)	(.221)
Lagged organizational growth	.036*	-.033*	.041*	-.027*
	(.004)	(.009)	(.004)	(.010)
Infant age range	.127	.075	.113	.095
	(.177)	(.216)	(.176)	(.215)
Toddler age range	.746*	-.341*	.765*	-.361*
	(.150)	(.192)	(.150)	(.191)

TABLE 3: Maximum Likelihood Models of DCC Organizational Niche Change[a] (Continued)

Independent variables	1 Expansion[b]	2 Contraction[c]	3 Expansion[b]	4 Contraction[c]
Preschool age range	-.849*	-.372*	-.832*	-.347*
	(.157)	(.195)	(.158)	(.198)
School-age range	-1.946*	-1.779*	-2.235*	-1.230*
	(.643)	(.437)	(.632)	(.422)
Purchase-of-service agreement (POSA)	1.008*	-1.103*	.804*	-.631*
	(.249)	(.381)	(.246)	(.345)
Site-sharing arrangement (SSA)	8.880*	-3.139*	8.432*	-1.320*
	(1.372)	(.899)	(1.465)	(.733)
Nonprofit	1.145*	-5.210*	.262	-3.106*
	(.584)	(.874)	(.572)	(.587)
Left-censored	-.109	.017	-.076	-.016
	(.234)	(.272)	(.234)	(.274)
Bank rate	-.033	.030	.002	.021
	(.037)	(.040)	(.037)	(.037)
Children's services division budget/ 100,000	.037	.008	.042*	.037
	(.026)	(.030)	(.025)	(.025)
Early childhood education graduates/ 100	.108	-.166	.116	-.174
	(.128)	(.147)	(.129)	(.145)
Organizational niche demand/ 1,000	-.031*	-.026*		
	(.008)	(.014)		
Other market segment demand/ 1,000	.020*	-.021*		
	(.006)	(.011)		
Local organizational niche demand/ 1,000			-.038*	-.057*
			(.016)	(.028)
Local other market segment demand/ 1,000			.044	.030
			(.029)	(.023)
Constant	-7.696*	-3.316	-1.071	-5.471*
	(1.690)	(2.234)	(1.813)	(1.477)
χ^2	641.01	545.02	649.69	551.57
Degrees of freedom	22	22	24	24

[a] Standard errors are in parentheses.
[b] The sample for the analysis of organizational niche expansion contained 6,402 annual spells and 277 organizational niche expansions.
[c] The sample for the analysis of organizational niche contraction contained 3,335 annual spells and 147 organizational niche contractions.

* $p < .05$

counteract the direct effects of organizational niche changes on the future probability of the same change. For example, while organizational niche contraction lowers the likelihood of future contraction, the significant negative coefficients for the age range dummies means that dropping any age range increases the likelihood. Indeed, the relative sizes of the coefficients indicate that if an initial contraction involved the school-age range, the future rate of contraction actually increased. Estimates for the environmental variables indicate that increases in the Children's Services Division budget increased the rate of organizational niche expansion. In addition, high demand in a DCC's current organizational niche lowered rates of organizational niche expansion and contraction and high demand in other market segments encouraged DCCs to enter them.

Overall, the estimates in Table 3 support our expectations that increases in overlap density would increase the rate of organizational niche expansion but not contraction (hypothesis 1) and that increases in nonoverlap density would lower the rate of both niche expansion and niche contraction (hypothesis 2).

ORGANIZATIONAL NICHE CHANGE AND FAILURE

Table 4 reports maximum-likelihood estimates for the analysis of DCC failure. Models 5 and 6 report baseline models of the effects of organizational niche change on DCC failure that include the effects of overlap density, nonoverlap density, and the organizational and environmental control variables. Model 6 also incorporates geographic distinctions. For convenience, estimates for overlap density, nonoverlap density, and the control variables are reported separately in Appendix E.

Preliminary support for structural inertia arguments (i.e., hypotheses 3, 4, and 5) is mixed. In support of hypothesis 3, the significant and positive estimate for organizational niche contraction and the significant and negative estimate for the organizational niche contraction time-clock indicate that the DCC failure rate jumps immediately after organizational niche contraction and then declines with time after the change. However, the organizational niche expansion and time-clock variables are not significant, failing to support hypothesis 3. Supporting hypothesis 4, the significant positive coefficient for the organizational niche contraction × age interaction means that organizational niche contraction increases the risk of failure with age. Supporting hypothesis 5, the significant negative coefficient for the organizational niche contraction × size interaction means that organizational niche contraction is less hazardous for large DCCs. In contrast to hypotheses 4 and 5, the coefficient for the organizational niche expansion × age interaction is significant and negative and the coefficient for the organizational niche expansion × size interaction is significant and positive. Thus, younger DCCs reap the benefits of organizational niche expansion more slowly and larger DCCs suffer a greater liability from change.

The complex implications of organizational niche contraction for DCC failure are illustrated graphically in Figure 1. The figure compares failure rates for DCCs that contract their organizational niche after one, five, and ten years of operation to failure rates of DCCs that do not make any change.[7] In the figure, a multiplier of greater (less) than 1 indicates that the failure rate is

TABLE 4: Maximum Likelihood Models of DCC Failure[a]

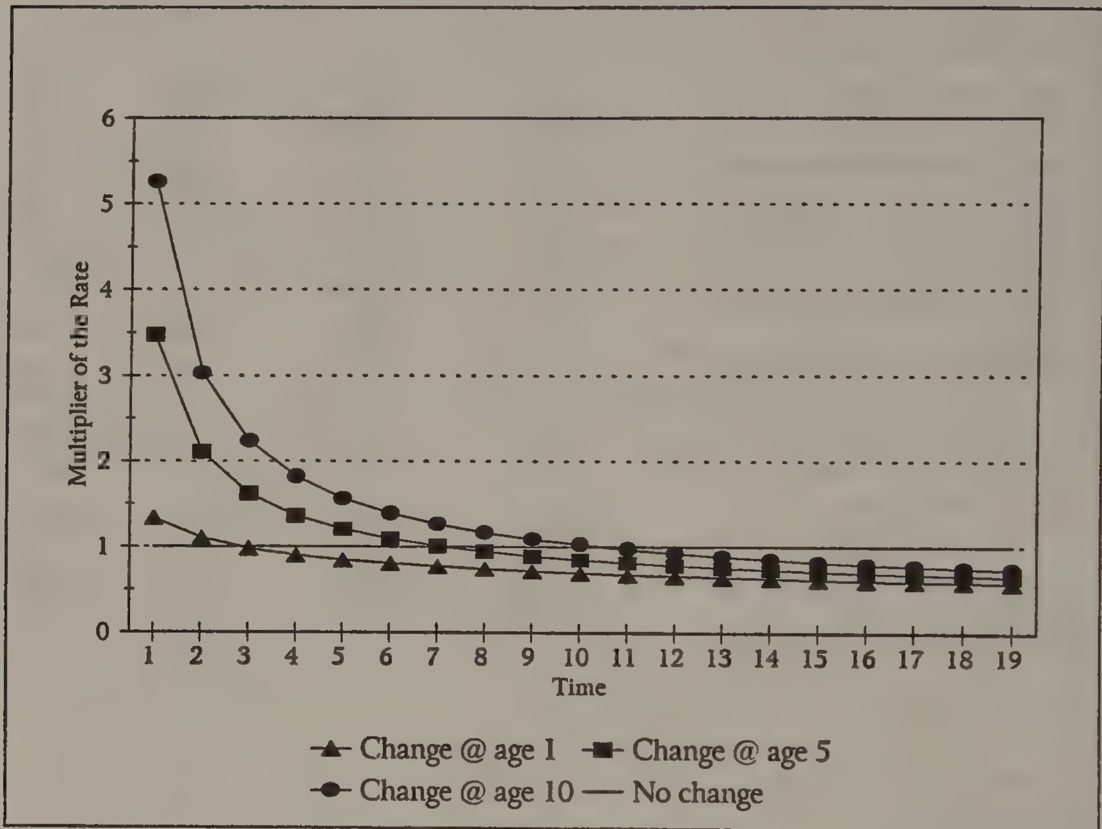
Independent variables	5	6	7	8
Organizational niche expansion	-1.166 (1.350)	-1.262 (1.352)	-.637 (1.411)	-.775 (1.421)
Organizational niche expansion time-clock	.322 (.368)	.340 (.365)	.212 (.371)	.255 (.368)
Organizational niche expansion x ln(age)	-.750* (.315)	-.765* (.317)	-.740* (.314)	-.761* (.317)
Organizational niche expansion x ln(size)	.588* (.310)	.547* (.313)	.539* (.319)	.528* (.322)
Organizational niche contraction	1.763* (.730)	1.835* (.738)	1.520* (.777)	1.484* (.784)
Organizational niche contraction time-clock	-.817* (.336)	-.803* (.333)	-.877* (.350)	-.879* (.349)
Organizational niche contraction x ln(age)	.623* (.330)	.622* (.330)	.607* (.335)	.597* (.337)
Organizational niche contraction x ln(size)	-.511* (.296)	-.541* (.301)	-.740* (.314)	-.732* (.323)
Δ Overlap density (expansion)			.003 (.005)	
Δ Overlap density (contraction)			.009* (.005)	
Δ Diffuse overlap density (expansion)				.004 (.009)
Δ Diffuse overlap density (contraction)				.008 (.007)
Δ Local overlap density (expansion)				.024* (.011)
Δ Local overlap density (contraction)				.037* (.020)
Constant	1.098 (1.329)	.272 (1.374)	1.210 (1.331)	.392 (1.377)
χ^2	179.42	187.74	184.13	194.66
Degrees of freedom	27	29	29	33

[a] Standard errors in parentheses; the sample contained 6,626 annual spells and 176 failures.

* $p < .05$

increased (decreased) relative to DCCs that have not changed by a factor equal to the multiplier. The multipliers in the figure show that, immediately after a change, failure rates for DCCs that contract organizational niches are up to five times higher than those that do not. The multipliers also show that the size of the initial jump increases with the age at the time of the change. However, the multipliers decline with time after the change and eventually fall below those for DCCs that do not change. A benefit is realized from the contraction after

FIGURE 1: Contraction and Failure

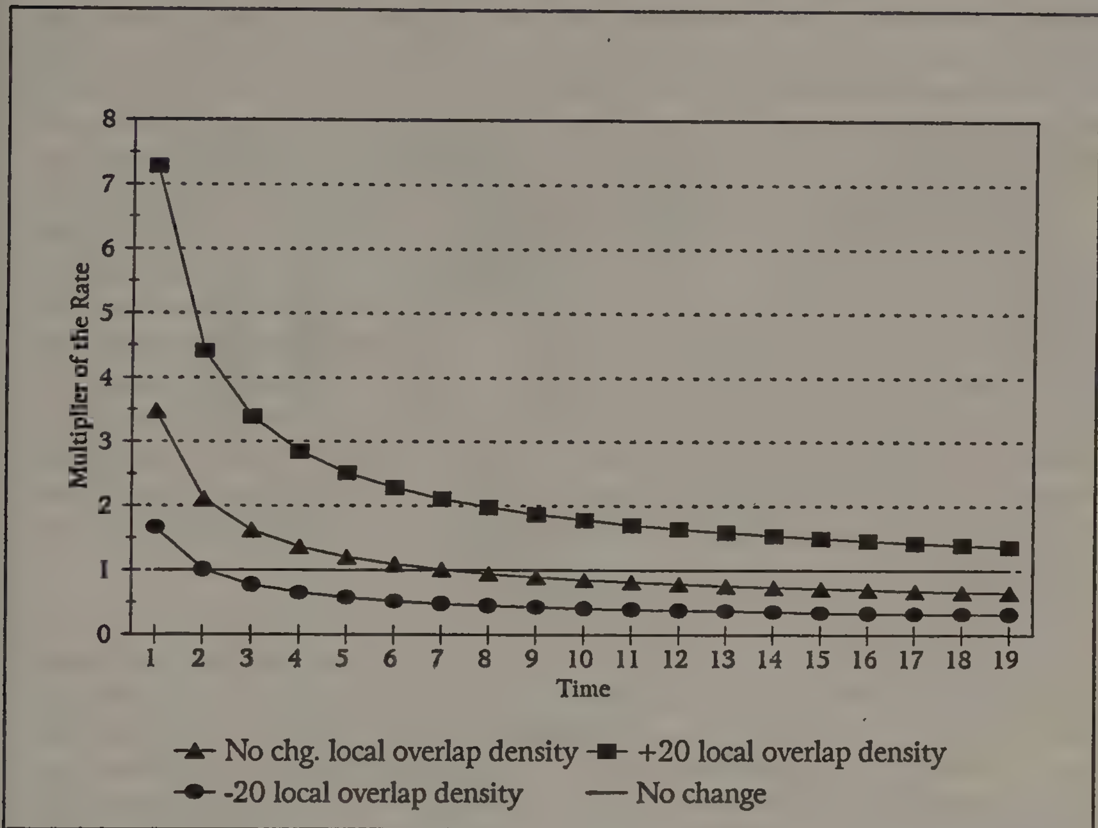


about three years for DCCs making the change at age one, and increases with the age of the DCC at the time of the change.

The effects of organizational niche changes on the intensity of competition are added in models 7 and 8. The significant positive coefficient for Δ overlap density (contraction) in model 7 indicates that organizational niche contractions that raised overlap density caused the failure rate to increase. However, Δ overlap density (expansion) is not significant.[8] In model 8, which incorporates the geographic disaggregation, the coefficients for Δ local overlap density resulting from organizational niche expansion and contraction are both significant and positive.[9] This means that DCCs that reduced competition with nearby DCCs by altering their organizational niches significantly improved their survival chances. Thus, as predicted in hypothesis 5, the survival consequences of organizational niche change depends on its effect on the intensity of competition, although the results are stronger at the constituent city level of analysis.

Combined with the coefficients for the immediate and dynamic effects of organizational niche changes, which remain unchanged in these models, the results indicate that by changing to organizational niches where competition was less intense, DCCs could increase the benefit of organizational niche expansion and moderate the disruptive effect of organizational niche contrac-

FIGURE 2: Contraction and Competition



tion. Figure 2 shows how moving to organizational niches where competition is lower moderates the disruptive effect of organizational niche contraction. This figure compares failure rates for DCCs that contract their organizational niche at age five in ways that leave local overlap density unchanged, increased by 20, and lowered by 20 to failure rates of DCCs that do not make any change. As the figure shows, compared to DCCs whose organizational niche contractions do not change local overlap density, DCCs that lower local overlap density by 20 reduce the magnitude of the initial jump in the failure rate by more than 50% and reduce the waiting time for realizing a survival advantage from seven (Figure 1) to two years. Clearly, DCCs that increase local overlap density by 20 face a quite different fate as a result of their actions.

Several of the control variables also significantly influence DCC failure rates (see Appendix E). While estimates for organizational age are insignificant, the failure rate exhibited negative size dependence. The estimates also indicate that DCCs with POSAs and SSAs and nonprofit DCCs had significantly lower failure rates. The estimates for the environmental control measures indicate that larger Children's Services Division budgets and greater organizational niche demand were associated with significant reductions in the DCC failure rate. The significant negative estimates for the lagged number of failures in the same

organizational niche as a focal DCC support Delacroix and Carroll's (1983) prediction that resources freed up by prior failures enhance the viability of surviving organizations. Finally, increases in overlap density and nonoverlap density respectively raise and lower the failure rate. The effects of local and diffuse overlap and nonoverlap densities are broadly parallel. Again consistent with our expectation that interactions among DCCs would be more intense at the local level, the coefficient estimates for local overlap density and local nonoverlap density are larger than the estimates for their diffuse counterparts.

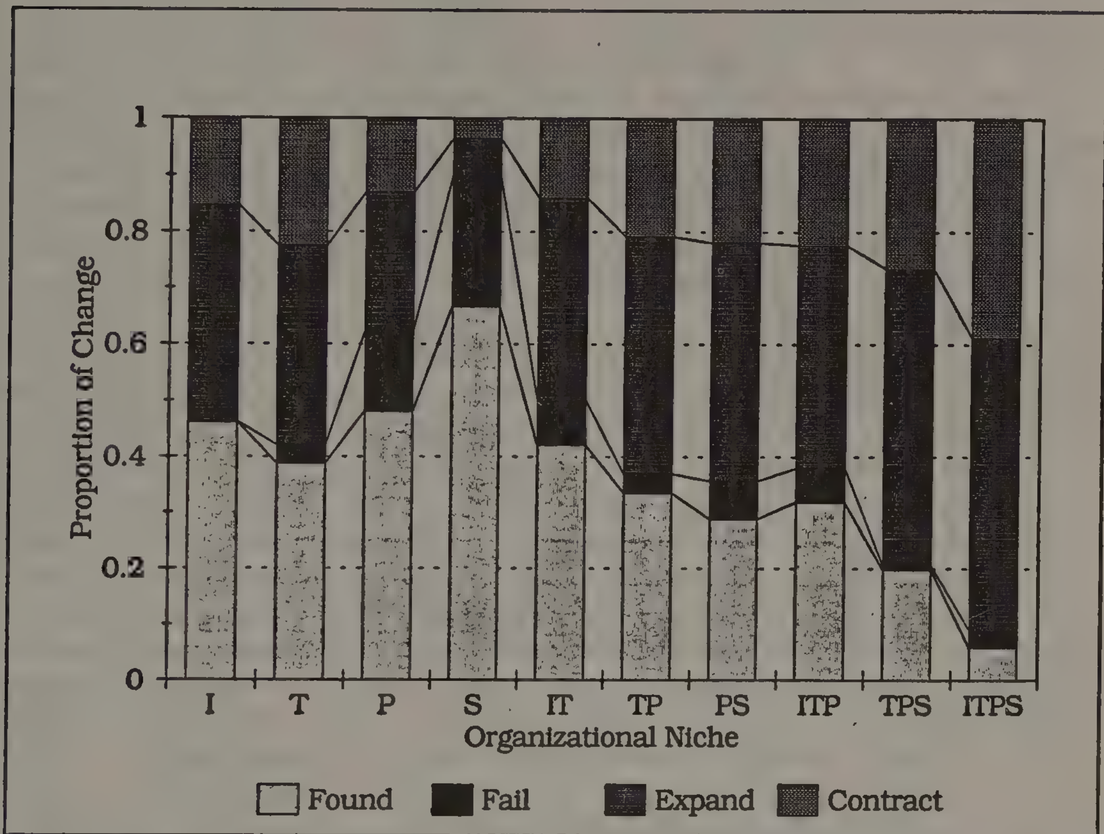
Overall, the estimates for organizational niche contraction in Table 4 support the predictions based on structural inertia theory that organizational niche changes would expose DCCs to an increased failure rate in the short run but that the rate would decline with time (hypothesis 3), and that older and smaller DCCs would be more likely to fail as a result of organizational niche changes (hypotheses 4 and 5). However, the estimates for organizational niche expansion did not support these hypotheses. In addition, the estimates for both niche expansion and niche contraction support our expectation that changes that increase overlap density (i.e., increase the intensity of competition) would increase the failure rate (hypothesis 6) — especially at the local level of analysis.

POPULATION-LEVEL IMPLICATIONS OF CHANGE AND FAILURE

To this point, we have focused on the implications of our analysis for change and failure at the level of individual DCCs. What remains unexplored, however, is how these processes combine to influence the evolution of the DCC population as a whole. To examine this question, we created Figure 3, which shows the contributions of organizational founding and failure and organizational niche expansion and contraction to the evolution of the DCC population between 1971 and 1989. The figure gives the proportions of entries into and exits from each organizational niche during the observation period accounted for by each type of change. For example, in the preschool (P) organizational niche, roughly 50% of all changes were foundings, 20% were organizational niche expansions, and 15% (each) were failures and organizational niche contractions.

Perhaps the most conspicuous feature of the figure is the predominance of foundings over other types of change, reflecting the growth of the DCC population from 1971 to 1989. Organizational niche expansion is the next most prevalent form of change, with organizational niche contraction and failure each providing much smaller contributions to the transformation of the population. It is also apparent that the contributions of the each change to population-level change varies widely — and systematically — across organizational niches. One particularly notable feature of the figure is that proportions of organizational foundings and failures on the one hand, and organizational niche expansions and contractions on the other hand, are each positively correlated with each other ($r = .71$ and $r = .87$, respectively). This implies that processes of transformation and selection tend to predominate in particular segments of the population. In this regard, while the proportions of organizational foundings and failures fall as the number of age ranges in an organizational niche increases ($r = -.87$ and $r = -.42$, respectively), the proportions of organizational niche expansions

FIGURE 3: Population-Level Change



and contractions grow with the number of age ranges ($r = .72$ and $r = .81$, respectively). Thus, in the DCC population, organizational transformation and selection do not characterize different regions of the macro-niche *per se*, but rather, transformation is much more likely to produce change among generalist organizational niches while birth and death are more likely to produce change among specialist organizational niches.

Beyond these relative influences of transformation and selection across the organizational niches making up the population, is it possible to conclude that one process contributes more than the other to the evolution of the population? Three considerations make it difficult to draw such a conclusion. First, although foundings lead numerically, many of these events contributed to the growth of the DCC population rather than to its changing demographic character. Second, because they result in simultaneous changes in two organizational niches, organizational niche expansions and contractions have twice the impact of organizational foundings and failures on population-level change. Third, and finally, because organizational niche changes affect the risk of failure, the combined population-level result is not the simple aggregate of each process separately, and is thus more complex than what is assumed in Figure 3.

Discussion and Conclusions

The purpose of this article was to investigate the roles of adaptation and selection processes in the evolution of organizational populations, a question central to contemporary organization theory. In this research we highlighted interorganizational competition as one primary force shaping patterns of organizational change and survival. The findings extend earlier research on adaptation and selection by (1) demonstrating how competition influences rates of organizational change, (2) showing how the survival consequences of organizational change can depend on its influence on competition, and (3) testing the full structural inertia theory specification of the survival consequences of organizational change.

The results provide evidence that the evolution of the DCC population was shaped jointly by processes of selection and adaptation. In support of the selectionist perspective, the survival prospects of DCCs were subject to the effects of competition and general environmental conditions. Thus, some DCCs responded to competition by dying. In addition, in the case of organizational niche contraction, structural inertia theory predictions about the survival consequences of change were supported fully. And, in the case of organizational niche expansion, although the findings did not support structural inertia theory, younger and larger DCCs still experienced liabilities of change. Lastly, while DCCs did not in general suffer from structural inertia (e.g., rates of change were not generally lower for older and larger DCCs), perhaps as a result of their general smallness and simple structures, DCCs did exhibit several constraints on their capacity for change (e.g., the damping effects of organizational niche expansion on its future probability and the rigidity of market contraction under competitive threat). In support of adaptationist perspectives, however, DCCs demonstrated a capacity to change their organizational niches, often in response to changes in competitive and environmental conditions, and without always jeopardizing their survival, although this was more pronounced in the case of organizational niche expansion than in that of contraction. Moreover, by moving to organizational niches with a lower intensity of local competition, DCCs could improve their survival chances significantly. Together, these findings suggest a complex relationship between adaptation and selection: because the relationship between organizational niche change and failure is moderated by the effects of organizational niche change on the intensity of competition, the population-level result of combined organizational niche change and competitive selection is not the simple aggregate of the separate processes.

Even though DCCs demonstrated a capacity to change organizational niches in ways that improved their survival chances significantly, this does not necessarily imply rational strategic calculation (Swaminathan & Delacroix 1991). The same result can be achieved on a probabilistic basis through random change attempts (March & Olsen 1976). Consequently, we think it is important to pursue further the question of the adaptiveness of DCCs. One way to explore this issue further is to examine whether DCCs moved to organizational niches where competition was, on average, less intense. In fact, the mean change in local overlap density resulting from both organizational niche expansions and contractions is essentially zero (see Appendix D). Thus, DCCs frequently

underwent organizational niche change, often in response to changing competitive conditions, but, on average, the changes DCCs made did not alter the intensity of competition they faced.

One possible interpretation for this absence of systematic effects of organizational niche change on the intensity of competition may be anchored in the bounded rationality of decision makers in organizations (March & Simon 1958). In the case of organizational niche expansion, it seems plausible that decision makers are more knowledgeable about the nature of competition they face in their current organizational niche from past experience than they are about competition in other organizational niches. And expanding their organizational niches would expose DCCs to a different set of competitors. Thus, intentionally rational expansions into new market segments may sometimes have the surprising consequence of exposing the organization to even greater competition, if the new organizational niches are densely populated. Of course, at other times, expansion into new segments may lower competition ex post facto. But, on average, there would be no effect. In the case of organizational niche contraction, since decision makers arguably possess superior knowledge of their current organizational niches, the absence of a systematic effect of organizational niche contraction on overlap density suggests that decision makers may not understand very well the competitive consequences of changing their organizational niches, i.e., that the effects of organizational niche change on the intensity of competition are a joint function of the resulting changes in resource overlap with potential competitors and organizational niche width.

Nevertheless, our results indicate that if DCCs moved to less competitive organizational niches they could improve their survival chances significantly. Thus, in the DCC population, the potential for adaptation, in the sense of improved longevity, is clear. The problem for DCCs is not that adaptive response is impossible or even rare, it is that decision makers are not, on average, able to distinguish between beneficial and detrimental organizational actions. Consequently, while our study provides clear evidence of the adaptive potential of DCCs, it offers little systematic support for a belief in the fundamental adaptability of organizations. Why isn't change more adaptive in the DCC population? One plausible explanation is that DCCs are incredibly nonbureaucratic organizations. This can be viewed as both an advantage (change is relatively easy), and a disadvantage (the ability to collect information on the environment is weak). The combination of reduced information gathering ability with a relatively organic organizational structure could account for our findings — a good deal of change, but not systematically effective change. Contained in this interpretation is a paradox: Organizations that can collect a lot of information often have organizational structures that are difficult to change, while those that can change the fastest often have trouble collecting relevant information to guide changes. Thus, although the organizational simplicity of DCCs might lead one to expect them to possess a high potential for developing *adaptive* responses to competition, in fact, it may lead them only to be responsive.[10] In this regard, it is interesting to note that the findings of this study broadly parallel those of Delacroix and Swaminathan's (1991) study of change among California wineries, a very different setting, but similarly unbureaucratic organizations. More generally, whether changes in response to environments

have adaptive consequences may depend upon the nature of learning processes in organizations (Levinthal 1991).

Several implications and directions for future research follow from this study. First, replications examining our hypotheses regarding competition and organizational change in other populations and for other kinds of changes are needed. Such replications would help clarify the role of organizational change in shaping population dynamics and help establish the generalizability and broader significance of the adaptive effects of competition-reducing changes for understanding the evolution of organizational populations. Implicit in this suggestion is the possibility that the dynamics of population-level change may be population-specific. Population-level change can occur through different rates of founding, failure, and change, and it may be that the prominence of these three mechanisms varies by population. Consequently, this study's results may be peculiar to the changes studied and, as implied above, the DCC population itself, which exists in a highly institutionalized context and comprises small organizations with ambiguous technologies. Future research aimed at identifying factors influencing the adaptiveness of change (e.g., information-gathering ability) may shed light on this question.

Second, distinguishing organizational niche changes by their effect on competition revealed hidden effects of these changes on failure. This may help account for earlier mixed results in studies examining the adaptive consequences of organizational change. In prior research, all instances of a particular kind of change were considered equivalent. While this may provide a reasonable approximation for some kinds of change (e.g., changes that have only intra-organizational implications), the findings of this study suggest that future research on the adaptiveness of organizational change needs to carefully consider the effect of changes on the intensity of competition. Third, our research highlights the importance of including both successful and unsuccessful organizations in studies of organizational change to avoid sample selection bias (Berk 1983; Heckman 1979). If we had studied only surviving firms, our conclusions about the effects of organizational change on performance would certainly have been quite different. For example, since DCCs that move to organizational niches where competition is less intense increase their survival chances, DCCs that move to organizational niches with lower competition are more likely to be survivors. Consequently, if we study only survivors, naturally changes will appear adaptive.

Fourth, in earlier research (Baum & Singh 1994a, 1994b) we studied the effects of overlap and nonoverlap densities on organizational founding and failure in the DCC population. This study extends the scope of our earlier work by examining the interrelated dynamics of overlap density, organizational change, and organizational failure. The results of all three studies support the characterization of interorganizational competition and mutualism implemented here, demonstrating the usefulness of examining the organizational niches of a population's members for understanding the dynamics of organizational populations. At the broadest level, we think the results of this study underscore the utility of examining how intrapopulation processes of competition and mutualism on the one hand, and selection and adaptation on the other hand, interweave to shape organizational evolution.

Notes

1. A counterintuitive though crucial point is that organizational niche expansions (contractions) do not necessarily lead to increases (decreases) in overlap density. This is because the effects of organizational niche changes on overlap density are a joint function of changes in organizational niche width and resource overlaps with potential competitors. We discuss this in detail below.

2. Appendix A shows the range of ages of children, in months, that can be enrolled by DCCs in each organizational niche and the matrix of overlaps among them (i.e., the portion of the client resource space that DCCs in the organizational niches occupy jointly).

3. For example, to qualify for a license to serve ten infants, in addition to planning and implementing activity and nutrition programs geared to the developmental needs of the children, a DCC must provide, among other things, three qualified staff members (i.e., diploma in early childhood education from a recognized institution), one separate activity space and one sleeping area that is separated from any other activity space, ten cradles or cribs and bedding, one table or counter space adjacent to a sink suitable for dressing or changing diapers, one flush toilet or hospital service sink, and 600 square feet of outdoor play space adjacent to the premises and fenced to a minimum height of four feet (Ontario Ministry of Community and Social Services 1988).

4. Metropolitan Toronto contains six cities: East York, Etobicoke, North York, Scarborough, Toronto, and York.

5. Appendix D present means, standard deviations, and bivariate correlations for all independent variables at metropolitan Toronto and local city levels of analysis, respectively. Although the correlations are generally significant, multicollinearity does not appear to be a problem.

6. While yearly data on working women and single parents were not available for every constituent city, census data (Social Planning Council of Metropolitan Toronto 1981) indicate that variation within metropolitan Toronto was not large. In 1981 the percentage of single-parent families with children from birth to five years of age ranged from 9.8 to 13.1%. In the same year, the percentage of women working full-time ranged from 58.1 to 61.5%.

7. Figure 1 is based on coefficients in model 6. The implications are not changed substantively in later models.

8. Although model 7 does not provide a significant improvement over model 5 ($\Delta\chi^2 = 4.71$, 2 df, n.s.), a reduced version of model 7 in which Δ overlap density (expansion) is removed provides a significant improvement ($\Delta\chi^2 = 4.22$, 1 df, $p < .05$) and the estimate for Δ overlap density (contraction) is not changed substantively.

9. Although model 8 does not provide a significant improvement over model 6 ($\Delta\chi^2 = 6.92$, 4 df, n.s.), a reduced version of model 8 in which Δ diffuse overlap density (expansion) and Δ diffuse overlap density (contraction) are removed provides a significant improvement ($\Delta\chi^2 = 6.48$, 2 df, $p < .05$) and the estimates for Δ local overlap density (expansion) and Δ local overlap density (contraction) are not changed substantively.

10. We are grateful to an anonymous reviewer for pointing this paradox out to us.

APPENDIX A: DCC Organizational Niche Resource Base Widths and Overlaps[a]

Org. Niche[b]	Organizational Niche[b] I	T	P	S	IT	TP	PS	ITP	TPS	ITPS
I	18									
T	0	12								
P	0	0	30							
S	0	0	0	60						
IT	18	12	0	0	30					
TP	0	12	30	0	12	42				
PS	0	0	30	60	0	30	90			
ITP	18	12	30	0	30	42	30	60		
TPS	0	12	30	60	12	42	90	42	102	
ITPS	18	12	30	60	30	42	90	60	102	120

[a] (1) The width (in months) of each age range is given on the diagonal;
(2) The overlaps (in months) between age ranges are reported below the diagonal;
(3) Only observed configurations of licensed age ranges are included;
(4) The widths (in months) of the age ranges are used to compute the raw resource overlaps, since the distribution of resources (i.e., children in Metropolitan Toronto) across the age ranges is roughly equal to the widths of the age ranges.

[b] Organizational niches:

I = Infant, 0-18 months
T = Toddler, 19-30 months
P = Preschool, 31-60 months
S = School-age, 61-120 months
IT = 0-30 months
TP = 19-60 months
PS = 31-120 months
ITP = 0-60 months
TPS = 19-120 months
ITPS = 0-120 months.

APPENDIX B: Resource Overlap Weights by Organizational Niche[a]

Org. Niche j[b]	Organizational Niche i[b] I	T	P	S	IT	TP	PS	ITP	TPS	ITPS
I	1	0	0	0	600	0	0	.300	0	.150
T	0	1	0	0	.400	.285	0	.200	.117	.100
P	0	0	1	0	0	.714	.333	.500	.294	.250
S	0	0	0	1	0	0	.667	0	.588	.500
IT	1	1	0	0	1	.285	0	.500	.117	.250
TP	0	1	1	0	.400	1	.333	.700	.411	.350
PS	0	0	1	1	0	.714	1	.500	.882	.750
ITP	1	1	1	0	1	1	.333	1	.411	.500
TPS	0	1	1	1	.400	1	1	.700	1	.850
ITPS	1	1	1	1	1	1	1	1	1	1

[a] The resource overlap weights w_{ij} in the table are the ratio of the overlaps between organizational niches i and j and the width of resource base of i. The values for the overlaps and widths are given in Table 1.

[b] See notes to Appendix C.

APPENDIX C: Overlap Densities by Organizational Niche, 1971-89[a]

Year	Organizational Niche[b] I	T	P	S	IT	TP	PS	ITP	TPS	ITPS
1971	9	14	104	58	10	77	73	57	66	57
1972	14	20	110	61	14	83	77	62	70	61
1973	21	26	120	70	21	91	86	70	79	70
1974	27	24	137	77	27	105	96	82	88	79
1975	33	36	164	78	34	127	106	99	98	88
1976	36	40	197	109	37	152	138	117	126	113
1977	36	37	214	120	36	163	151	125	137	122
1978	39	40	222	120	41	171	153	131	140	125
1979	33	45	238	142	37	182	173	137	158	139
1980	35	51	257	163	41	197	194	149	177	156
1981	33	53	268	177	41	206	207	154	189	165
1982	37	62	301	198	47	232	232	174	212	186
1983	41	72	319	211	53	247	246	185	226	198
1984	43	85	339	221	59	265	260	199	239	210
1985	53	120	370	234	78	297	279	224	259	229
1986	67	151	415	264	99	338	314	257	294	260
1987	84	187	459	288	125	381	344	292	326	290
1988	88	199	474	300	131	394	357	302	338	301
1989	97	192	481	299	141	402	359	311	341	305

[a] Computations described in text. All values are rounded to the nearest number of organizations.

[b] Organizational niches:

I = Infant, 0-18 months
T = Toddler, 19-30 months
P = Preschool, 31-60 months
S = School-age, 61-120 months
IT = 0-30 months
TP = 19-60 months
PS = 31-120 months
ITP = 0-60 months
TPS = 19-120 months
ITPS = 0-120 months.

APPENDIX D: Means, Standard Deviations, and Pearson Correlations for Independent Variables[a]

Panel A: Metropolitan Toronto Level of Analysis

Variable	Mean	(S.D.)	1	2	3	4
1. Overlap density	244.0	(103.7)	1.000			
2. Nonoverlap density	152.1	(89.1)	.195	1.000		
3. Δ overlap density (expansion)	-.044	(22.9)	-.056	-.031	1.000	
4. Δ overlap density (contraction)	.363	(27.5)	.189	-.132	-.181	1.000
5. Cumulative organizational niche expansion	.172	(.377)	.103	.133	-.227	.174
6. Cumulative organizational niche contraction	.151	(.358)	.083	-.067	-.168	.306
7. Org. niche expansion clock	.273	(.651)	.137	.117	.252	.189
8. Org. niche contraction clock	.269	(.682)	.121	-.016	-.069	.222
9. Org. niche expansion × ln(age)	.428	(.964)	.094	.127	-.264	-.177
10. Org. niche contraction × ln(age)	.378	(.918)	.070	-.061	-.179	.239
11. Org. niche expansion × ln(size)	.670	(1.51)	.101	.142	-.278	.181
12. Org. niche contraction × ln(size)	.611	(1.48)	.074	-.049	.164	-.282

Panel B. Local City Level of Analysis

Variable	Mean	(S.D.)	1	2	3	4
1. Diffuse overlap density	159.4	(68.2)	1.000			
2. Local overlap density	84.5	(35.9)	.782	1.000		
3. Diffuse nonoverlap density	99.7	(59.6)	.246	.202	1.000	
4. Local nonoverlap density	52.4	(29.8)	.137	.083	.784	1.000
5. Δ Diffuse overlap density (expansion)	.298	(17.0)	-.060	-.055	-.033	-.031
6. Δ Local overlap density (expansion)	-.342	(6.98)	-.041	-.037	-.023	-.021
7. Δ Diffuse overlap density (contraction)	.384	(21.6)	.183	.199	-.119	-.152
8. Δ Local overlap density (contraction)	-.021	(7.39)	.145	.159	-.103	-.130
9. Cumulative organizational niche expansion			.105	.097	.134	.125
10. Cumulative organizational niche contraction			.077	.092	-.060	-.080
11. Org. niche expansion clock			.138	.134	.121	.107
12. Org. niche contraction clock			.115	.129	-.008	-.031
13. Org. niche expansion × ln(age)			.097	.089	.129	.121
14. Org. niche contraction × ln(age)			.066	.078	-.056	-.072
15. Org. niche expansion × ln(size)			.103	.094	.144	.136
16. Org. niche contraction × ln(size)			.068	.083	-.043	-.062

[a] The sample contained 6,626 annual spells. Correlations >.025 significant at $p < .05$.

APPENDIX D: Means, Standard Deviations, and Pearson Correlations for Independent Variables[a] (Continued)

Panel A: Metropolitan Toronto Level of Analysis

Variable	5	6	7	8	9	10	11
1. Overlap density							
2. Nonoverlap density							
3. Δ overlap density (expansion)							
4. Δ overlap density (contraction)							
5. Cumulative org. niche expansion	1.000						
6. Cumulative org. niche contraction	.224	1.000					
7. Org. niche expansion clock	.821	.248	1.000				
8. Org. niche contraction clock	.235	.834	.270	1.000			
9. Org. niche expansion × ln(age)	.802	.227	.716	.239	1.000		
10. Org. niche contraction × ln(age)	.220	.817	.241	.770	.278	1.000	
11. Org. niche expansion × ln(size)	.878	.225	.792	.233	.819	.239	1.000
12. Org. niche contraction × ln(size)	.307	.878	.324	.812	.329	.833	.236

Panel B: Local City Level of Analysis

Variable	5	6	7	8
1. Diffuse overlap density				
2. Local overlap density				
3. Diffuse nonoverlap density				
4. Local nonoverlap density				
5. Δ Diffuse overlap density (expansion)	1.000			
6. Δ Local overlap density (expansion)	.681	1.000		
7. Δ Diffuse overlap density (contraction)	-.175	-.137	1.000	
8. Δ Local overlap density (contraction)	-.144	-.207	.638	1.000
9. Cumulative org. niche expansion	-.276	-.206	.167	.139
10. Cumulative org. niche contraction	-.173	-.132	.397	.322
11. Org. niche expansion clock	.257	.184	.185	.141
12. Org. niche contraction clock	-.075	.027	.413	.324
13. Org. niche expansion × ln(age)	.258	.132	-.167	-.048
14. Org. niche contraction × ln(age)	-.181	-.146	.327	.248
15. Org. niche expansion × ln(size)	-.284	-.206	.077	.240
16. Org. niche contraction × ln(size)	.170	.123	-.277	-.360

[a] The sample contained 6,626 annual spells. Correlations > .025 significant at $p < .05$.

APPENDIX E: Control Variable Estimates for DCC Failure Models 5-8[a]

	5	6	7	8
ln(organizational age)	.193	.198	.190	.196
	(.129)	(.130)	(.128)	(.130)
ln(organizational size)	-.833*	-.801*	-.855*	-.820*
	(.158)	(.157)	(.159)	(.158)
Lagged organizational growth	-.011	-.009	-.013	-.013
	(.015)	(.015)	(.015)	(.015)
Purchase-of-service agreement (POSA)	-.377*	-.390*	-.347*	-.355*
	(.175)	(.176)	(.177)	(.177)
Site-sharing arrangement (SSA)	-1.031*	-1.019*	-1.032*	-1.027*
	(.187)	(.187)	(.188)	(.188)
Nonprofit orientation	-.352*	-.358*	-.338*	-.350*
	(.195)	(.195)	(.196)	(.197)
Infant age range	.841	1.135	.850	1.112
	(.945)	(.963)	(.945)	(.965)
Toddler age range	-.468	-.313	-.497	-.320
	(.844)	(.865)	(.847)	(.868)
Preschool age range	-.300	.641	-.281	.670
	(3.167)	(3.251)	(3.167)	(3.253)
School-age range	1.673	2.207	1.694	2.244
	(3.392)	(3.456)	(3.394)	(3.459)
Left-censored	-.113	-.052	-.153	-.083
	(.311)	(.310)	(.314)	(.314)
Bank rate	.103*	.105*	.103*	.105*
	(.044)	(.044)	(.044)	(.044)
Children's services division budget/100,000	-.085*	-.089*	-.083*	-.084*
	(.036)	(.047)	(.036)	(.047)
Early childhood educ. graduates/100	-.087	-.074	-.092	-.080
	(.157)	(.164)	(.157)	(.164)
Organizational niche demand/1,000	-.044*		-.045*	
	(.022)		(.022)	
Local organizational niche demand/1,000		-.034*		-.038*
		(.018)		(.018)
Org. niche foundings	.008		.007	
	(.012)		(.012)	
Organizational niche failures	-.104*		-.107*	
	(.046)		(.046)	
Local org. niche foundings		.010		.010
		(.013)		(.013)
Local org. niche failures		-.112*		-.115*
		(.048)		(.048)
Overlap density	.009*		.008*	
	(.004)		(.004)	
Nonoverlap density	-.013*		-.013*	
	(.004)		(.004)	
Diffuse overlap density		.011*		.010*
		(.004)		(.005)
Local overlap density		.024*		.022*
		(.009)		(.010)
Diffuse nonoverlap density		-.012*		-.012*
		(.005)		(.005)
Local nonoverlap density		-.018*		-.020*
		(.006)		(.006)

[a] Standard errors in parentheses; the sample contained 6,626 annual spells and 176 failures. * $p < .05$

References

Amburgey, Terry L. 1986. "Multivariate Point Processes in Social Research." *Social Science Research* 15:190-207.

Amburgey, Terry L., Dawn Kelly, and William P. Barnett. 1993. "Resetting the Clock: The Dynamics of Organizational Change and Failure." *Administrative Science Quarterly* 38:51-73.

Barnett, William P. 1990. "The Organizational Ecology of a Technological System." *Administrative Science Quarterly* 35:31-60.

Barnett, William P., and Glenn R. Carroll. 1987. "Competition and Mutualism among Early Telephone Companies." *Administrative Science Quarterly* 32:400-421.

Baum, Joel A.C. 1990. "Inertial and Adaptive Patterns in the Dynamics of Organizational Change." Pp. 165-69 in *Academy of Management Proceedings*, edited by J.L. Wall and L.R. Jauch. San Francisco, Cal.: Academy of Management.

____. 1996. "Organizational Ecology." Pp. 77-114 in *Handbook of Organizational Studies*, edited by Steward Clegg, Cynthia Hardy, and Walter Nord. Sage.

Baum, Joel A.C., and Stephen J. Mezias. 1992. "Localized Competition and Organizational Failure in the Manhattan Hotel Industry." *Administrative Science Quarterly* 37:580-604.

Baum, Joel A.C., and Christine Oliver. 1991. "Institutional Linkages and Organizational Mortality." *Administrative Science Quarterly* 36:187-218.

Baum, Joel A.C., and Jitendra V. Singh. 1994a. "Organizational Niche Overlap and the Dynamics of Organizational Mortality." *American Sociological Review* 100:346-80.

____. 1994b. "Organizational Niche Overlap and the Dynamics of Organizational Founding." *Organization Science* 5:483-501.

Berk, Richard A. 1983. "An Introduction to Sample Selection Bias in Sociological Data." *American Sociological Review* 48:386-99.

Bourgeois, Louis J. 1984. "Strategic Management and Determinism." *Academy of Management Review* 9:586-96.

Carroll, Glenn R., and Michael T. Hannan. 1989. "Density Delay in the Evolution of Organizational Populations: A Model and Five Empirical Tests." *Administrative Science Quarterly* 34:411-30.

Chamberlin, Edward H. 1933. *The Theory of Monopolistic Competition*. Harvard University Press.

Chandler, Alfred. 1977. *The Visible Hand*. Belknap.

Child, John. 1972. "Organization Structure, Environment and Performance: The Role of Strategic Choice." *Sociology* 6:1-22.

City of Toronto Planning and Development Department. 1983. "Social Change in Toronto: A Context for Human Services Planning." City of Toronto Planning and Development Department.

Community Information Center of Metropolitan Toronto. 1971-1989. *Directory of Day Care Centers and Nursery Schools in Metropolitan Toronto*. Community Information Center of Metropolitan Toronto.

Delacroix, Jacques, and Glenn R. Carroll. 1983. "Organizational Foundings: An Ecological Study of the Newspaper Industries of Argentina and Ireland." *Administrative Science Quarterly* 28:274-91.

Delacroix, Jacques, and Anand Swaminathan. 1991. "Cosmetic, Speculative, and Adaptive Organizational Change in the Wine Industry: A Longitudinal Study." *Administrative Science Quarterly* 36:631-61.

Delacroix, Jacques, Anand Swaminathan, and Michael E. Solt. 1989. "Density Dependence versus Population Dynamics: An Ecological Study of Failings in the California Wine Industry." *American Sociological Review* 54:245-62.

Fligstein, Neil, and Kenneth Dauber. 1989. "Structural Change in Corporate Organizations." Pp. 75-96 in *Annual Review of Sociology*. Vol. 15, edited by W. Richard Scott and Judith Blake. Annual Reviews.

Fombrun, Charles J. 1986. "Structural Dynamics within and between Organizations." *Administrative Science Quarterly* 31:403-21.

Ginsberg, Ari, and Joel A.C. Baum. 1994. "Evolutionary Processes and Patterns of Core Business Change." Pp. 127-51 in *Evolutionary Dynamics of Organizations*, edited by Joel A.C. Baum and Jitendra V. Singh. Oxford University Press.

Hannan, Michael T., and Glenn R. Carroll. 1992. *Dynamics of Organizational Populations: Density, Competition, and Legitimation*. Oxford University Press.

Hannan, Michael T., and John H. Freeman. 1977. "The Population Ecology of Organizations." *American Journal of Sociology* 83:929-84.

____. 1984. "Structural Inertia and Organizational Change." *American Sociological Review* 49:149-64.

____. 1989. *Organizational Ecology*. Harvard University Press.

Hannan, Michael T., James Ranger-Moore, and Jane Banaszak-Holl. 1990. "Competition and the Evolution of Organizational Size Distributions." Pp. 246-68 in *Organizational Evolution: New Directions*, edited by Jitendra V. Singh. Sage.

Haveman, Heather A. 1993. "Ghosts of Managers Past: Managerial Succession and Organizational Mortality." *Academy of Management Journal* 36:864-81.

____. 1994. "The Ecological Dynamics of Organizational Change: Density and Mass Dependence in Rates of Entry into New Markets." Pp. 152-66 in *Evolutionary Dynamics of Organizations*, edited by Joel A.C. Baum and Jitendra V. Singh. Oxford University Press.

Hawley, Amos. 1950. *Human Ecology: A Theory of Community Structure*. Ronald.

Heckman, James J. 1979. "Sample Selection Bias as Specification Error." *Econometrica* 47:153-61.

Hrebiniak, Lawrence G., and William F. Joyce. 1985. "Organizational Adaptation: Strategic Choice and Environmental Determinism." *Administrative Science Quarterly* 30:336-49.

Kelly, Dawn, and Terry L. Amburgey. 1991. "Organizational Inertia and Momentum: A Dynamic Model of Strategic Change." *Academy of Management Journal* 34:591-612.

Levinthal, Daniel A. 1991. "Organizational Adaptation and Environmental Selection — Interrelated Processes of Change." *Organization Science* 2:140-45.

March, James G., and Johan P. Olsen. 1976. *Ambiguity and Choice in Organizations*. Bergen, Norway: Universitetsförlaget.

March, James G., and Herbert A. Simon. 1958. *Organizations*. Wiley.

Maynard, Fredelle B. 1985. *The Child Care Crisis*. Viking.

McKelvey, Bill. 1982. *Organizational Systematics*. University of California Press.

McPherson, J. Miller. 1983. "An Ecology of Affiliation." *American Sociological Review* 48:519-32.

Miles, Raymond E., and Charles C. Snow. 1978. *Organizational Strategy, Structure, and Process*. McGraw-Hill.

Miner, Anne S., Terry L. Amburgey, and Timothy Stearns. 1990. "Interorganizational Linkages and Population Dynamics: Buffering and Transformational Shields." *Administrative Science Quarterly* 35:689-713.

Mitchell, Will. 1989. "Whether and When? Probability and Timing of Incumbents' Entry into Emerging Industrial Subfields." *Administrative Science Quarterly* 34:208-30.

Ontario Ministry of Community and Social Services. 1988. *Day Nurseries Act*. Ontario Ministry of Community and Social Services.

Pfeffer, Jeffrey, and Gerald R. Salancik. 1978. *The External Control of Organizations*. Harper & Row.

Porter, Michael E. 1980. *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. Free Press.

Ramanujam, Vasudevan, and P. Varadarajan. 1989. "Research on Corporate Diversification: A Synthesis." *Strategic Management Journal* 10:523-51.

Rumelt, Richard. 1986. *Strategy, Structure, and Economic Performance* Harvard University Press.

Scott, W. Richard. 1992. *Organizations: Rational, Natural and Open Systems*. 3d ed. Prentice-Hall.

Singh, Jitendra V., David J. Tucker, and Agnes G. Meinhard. 1988. "Are Voluntary Social Service Organizations Structurally Inert? Exploring an Assumption in Organizational Ecology." Paper presented at an Academy of Management Meeting, Anaheim, Cal.

____. 1991. "Institutional Change and Ecological Dynamics." Pp. 390-422 in *The New Institutionalism in Organizational Analysis*, edited by Walter W. Powell and Paul J. DiMaggio. University of Chicago Press.

Stinchcombe, Arthur L. 1965. "Organizations and Social Structure." Pp. 153-93 in *Handbook of Organizations*, edited by James G. March. Rand McNally.

Social Planning Council of Metropolitan Toronto. 1981. *Metropolitan Social Profile: Guide to Social Planning Facts in Metropolitan Toronto*. Vol. 1. Social Planning Council of Metropolitan Toronto.

Swaminathan, Anand, and Jacques Delacroix. 1991. "Differentiation within an Organizational Population: Additional Evidence from the Wine Industry." *Academy of Management Journal* 34:679-92.

Thompson, James D. 1967. *Organizations in Action*. Harper & Row.

Tuma, Nancy B. 1980. *Invoking Rate*. 2d ed. SRI International.

Tuma, Nancy B., and Michael T. Hannan. 1984. *Social Dynamics: Models and Methods*. Academic Press.

The Settlement Experience of Latinos in Chicago: Segregation, Speculation, and the Ecology Model

JOHN J. BETANCUR, *University of Illinois at Chicago*

Abstract

This article describes the process and characteristics of Latino settlement in Chicago. It identifies elements of segregation and discrimination associated with Latino status. The analysis disagrees with Massey's claim that Latino settlement is proceeding along the lines of the ecological model. It claims that, though descriptively relevant, the model is not very useful for explaining *the Latino experience, particularly prejudices and practices related to the status of Latinos as a dominated group in the U.S. The author claims that a combined historical, qualitative, and quantitative analysis reveals crucial qualifications that statistical analyses often miss.*

This article addresses Latino settlement in Chicago. It contends that this experience has been characterized by residential exclusion and speculation rather than by opportunity and mobility. The analysis disagrees with descriptions of Latino settlement based on the ecology model (Aguirre, Schwirian & LaGreca 1980; Massey 1987). It claims that the model, though descriptively relevant, is not very useful for *explaining* the Latino experience. The argument is specifically different from Massey's (1987) version, which contends that while the ecological model cannot explain black residential distribution, it is valid for Latinos. Statistical evidence may suggest the relevance of aspects of the model, but it misses key qualifications and elements suggested here.

The argument is constructed in four sections. The first argues that Latino immigration reflected U.S. domination of Latin America. The second examines ways in which the resulting vulnerability and status of Latinos allowed for real-estate speculation to exploit their quasi-racial status and highlights the resulting processes of residential exclusion. The third describes the settlement patterns of Latinos today and brings up additional evidence questioning the applicability of the ecological model to them. It concludes with a brief discussion of findings.

* *This article profited significantly from the comments and suggestions of David Ranney, Wim Wiewel, and three anonymous reviewers. Walter Palomino provided valuable assistance in the production of figures and statistics.* *Please direct correspondence to John J. Betancur, University of Illinois at Chicago, 400 S. Peoria, Suite 2100, Chicago, Illinois 60607. Tel. 312-996-4323, Fax 312-996-5766.*

 Social Forces, June 1996, 74(4):1299-1324

The Ecological Model

The ecological model either in the original form or in later versions has dominated analyses of residential segregation. The initial version explained urban change in terms of competition and succession among land uses. In residential areas, one ethnic group succeeded another one when the latter left. Alternatively, groups engaged in competition and displaced each other. Underlying this process was the assumption of mobility: as groups improved economically, they moved into better areas. Neighborhood change was, then, explained in terms of invasion, succession, and mobility (Park & Burgess 1921; Park, Burgess & McKenzie 1925). A more recent version of the ecological model, *social area analysis,* views residential differentiation as reflecting the increasing complexity associated with development. As development produces different groups, people of similar backgrounds seek to live with their kind and the result is social separation and segregation (Berry & Horton 1970; Berry & Kasarda 1977; Shevky & Bell 1955).

Along these lines, Massey (1987) proposed an updated version. In his view, ethnic segregation stems from "the interplay of the opposing forces of concentration and dispersion" (317). Concentration results from increasing spatial differentiation; dispersion, from mobility and acculturation. Immigration is a case in point. Immigrants hold low-wage jobs and need to minimize costs. Migration chains lead them to low-rental ethnic areas, near work, relatives, and friends. If differences with current residents are too deep, a process of replacement follows. When they "reach a threshold of density" (318), they develop their own social infrastructure, their enclaves become the basis for solidarity, and more members join the group (concentration). As they acquire the language, culture, and values of the host society, they get better jobs and disperse into higher-income native areas with better services. Since, by that time, they have developed the values and status as natives, their move is not likely to produce a new process of replacement and succession but one of integration and assimilation.

Massey's (1987) update provides a version that is more flexible, while getting rid of outdated elements ("a rigid biological analogy and a sentimental, ethnocentric view of ethnicity now discredited") and introducing others left out by previous formulations ("larger processes of social change and economic development"). He removes one of the most controversial applications of the model by leaving blacks out of the equation. Studies of segregation, he claims, "suggest that black segregation stems more from white prejudice than from the socioeconomic processes specified by the ecological model" (316). Meanwhile, he includes Latinos as "ethnics" properly represented by the model (326-28). This decision, we argue, misses the significance of the Latino status. Not only have Latinos been discriminated against because they are Latino, but discrimination has been reflected in unique forms of manipulation and exploitation affecting their settlement and mobility in ways that differ significantly from those of European immigrants.[1] Thus, while the literature refers to Latinos as ethnic, national, or racial groups, often without much attention to the implications of the chosen term, this article argues for the need to construct the term historical-

ly. For this, the article engages in analysis that emphasizes the common bonding of subjugation of Latinos in the U.S. and its specific residential manifestations. The terms *racialization* and *race*, for instance, are used to highlight the application of prejudice and discriminatory practices in the housing market to them because they are Latino. Generally, however, the article uses the term *Latino* within the context of domination established in the next section.

Latino Importation or the Production of a New Urban Minority

The migration of Latinos to the U.S. and Chicago has been closely associated with economic exploitation and political domination of Latin America by the U.S. (Bach 1978; Betancur, Córdova & Torres 1993; Bonilla & Campos 1981; Centro de Estudios Puertorriqueños 1979; Fernández Kelly 1983; Portes and Walton 1981; Sassen 1988; Teitelbaum 1985).[2] Along these lines, the experience of Latinos has been characterized by extreme forms of domination and exclusion because they are Latino (Acuña 1988; Pachón 1985; Padilla 1987). Employer abuses, labor segmentation and immobility, and segregation are clear manifestations of this (Barrera 1979; Blauner 1969; Bullock & Manus 1990; Lane & Escobar 1987; Meléndez 1988; Omi & Winant 1986; Padilla 1987; Reisler 1976; Taylor 1932; Yinger 1985).

The presence of Latinos in the U.S. is largely the result of forceful annexation or purchase of Latin American territory and people by the U.S. (northern Mexico, Puerto Rico); use of migration as an escape valve for explosive social conditions (Puerto Rico, Central America); a strategy of the cold war (Cuba, Nicaragua, Dominican Republic); distortions related to U.S. investment in Latin America (Sassen 1988); or a convenient, temporary solution to so-called labor shortages (Kirstein 1977). These actions have followed from U.S. claims on the region as its sphere of influence.[3]

Northern Mexico became part of the U.S. through conquest and purchase (1848). This is also true for Puerto Rico as a result of the U.S.-Spain colonial war of 1898. Migration to the U.S. resulted from the ensuing transformation of the island by U.S. interests (Sanchez-Korrol 1983). Dependent development in Mexico led to massive displacement of peasants who were then imported by U.S. employers and turned into a dominated segment of disposable, cheap labor. Not only were they admitted under special arrangements (e.g., the Bracero Program of 1942-64 and the Temporary Admissions Program of World War I), but employers and government colluded to generate a stream of undocumented and unprotected workers (Corwin 1978; Kirstein 1977; Kiser & Kiser 1979).

A steady and large flow of Mexicans to Chicago resulted from their *importation* by railroad, steel, and other industries,[4] as part of the industries' strategies of labor control (Año Nuevo Kerr 1976; Taylor 1932). These industries systematically subjected Mexicans to labor segmentation (Año Nuevo Kerr 1976; Lane & Escobar 1987; Taylor 1932). Large imports of temporary Mexican workers to Chicago and the Midwest took place in the periods 1916-28 and 1942-64. Many remained or came back as undocumented workers. Their

importation encouraged others to follow and became the basis for chain and circular migration ever since.[5]

Mexicans have been employed and laid off at convenience, imported, deported, and prevented from competing for better jobs.[6] The nativist debate of the 1920s portraying them as "perhaps the worst element that comes into the U.S.," "serf, peon and slave types," "undesirable people," and "inassimilable manpower" (Reisler 1976:200-205); the bitter repatriation drives of the 1921 and 1929 depressions; and the 1954 "wetback" and later operations of massive deportation speak to this. They testify to the framing of Mexicans as a cultural and racial threat, the justification of their use for unskilled physical work, and the claim that they were not worthy of full participation in U.S. society (Reisler 1976). Scapegoating of Mexicans has continued through the years.[7]

Most recently, they are being lured into Chicago by minimum-wage jobs in sweatshop manufacturing and services. Employers in need of unskilled workers depend largely on them. As a result of the terms of migration, the types of workers recruited, the jobs open to them, and chains tying previous entrants to new entrants, their immigration has been highly selective of poor, unskilled, uneducated Mexicans. These factors have led them to industries and occupations similar to those of their ancestors. The conditions and selectivity of their immigration led to communities with very little Mexican leadership or professional cadres.

The immigration of a large group of Puerto Ricans[8] to Chicago started with importation of contract labor for service, farm, and manufacturing work after World War II, with the cooperation of authorities at both ends.[9] These recruits became the nuclei and the impetus for further recruitment and chain migration. Workers were imported by employers unable to secure U.S. labor under the conditions and at the wages offered (Maldonado 1987; Padilla 1947). Once in Chicago, they faced highly exploitative conditions — wages below those prevailing for similar work; transfers without notice; long hours; deduction of recruiter fees and transportation costs from their paychecks; poor housing and lodging (Martínez 1989; Padilla 1947). Even though Puerto Ricans are U.S. citizens, the terms and conditions of their importation and employment have been close to those of Mexicans. They have been categorized and treated as blacks (Padilla 1987). Unlike European ethnics, the terms of their importation produced a largely single-class group without the cadres required for the development of healthy communities.

Today, after more than a century in the U.S., the white-Latino gap has not improved (Bean & Tienda 1987; Morales & Bonilla 1993). Though equal to whites in the eyes of the law, Latinos have been consistently unequal politically and socioeconomically (Hero 1992). As Barrera (1979) argues, their condition has been shaped through a combination of conquest, arranged immigration,[10] and other forms of domination amounting to internal colonization. Studies of both groups in the U.S. describe their experience as one of political domination and economic exploitation (Acuña 1988; García 1988; Jennings & Rivera 1984; Sánchez-Korrol 1983). In Chicago, the income and occupational gap between Latinos and whites has been increasing (Betancur, Córdova & Torres 1993).

Hence, Mexicans and Puerto Ricans came into the U.S. under conditions that were essentially different from those of Europeans on whom the ecology

model is based. Most Europeans did not undergo this process of manipulation and domination, received a permanent migrant status from the beginning, and were extended the opportunities of earlier white immigrants. In contrast, Latinos were shaped into an internally dominated labor segment, second-class citizens, indeed, a minority group largely along the lines of blacks. This condition has been reproduced over the years.

The Formation of Segregated Latino Areas

In Chicago, Mexicans were initially housed in bunkhouses or railroad camps or were directed to boarding houses by the companies (Taylor 1932).[11] This housing was generally temporary — as their job assignments were meant to be. With increases in their numbers and extension of their stay, they sought housing near the plants. Meanwhile, used to exploiting ethnic feelings, slumlords found a new source of profit — Mexican origin — among the newcomers.

The landmark research of Taylor, Abbott, Hughes, Constable, Jones, Britton, and Redfield is extremely valuable here. According to them, landlords rented to Mexicans vacant, unhealthy, run-down property in pockets of deterioration where Europeans would not live (Hughes 1925, 1937; Jones [1928] 1971; Redfield 1924-25; Taylor 1932). They ended up in back parts, rear houses, and basements (Briton & Constable quoted by Taylor 1932). Restriction to the poorest areas "largely among Negroes" (224), produced clusters of artificial scarcity that allowed landlords to charge Mexicans — and blacks — higher rents than they charged any other groups.

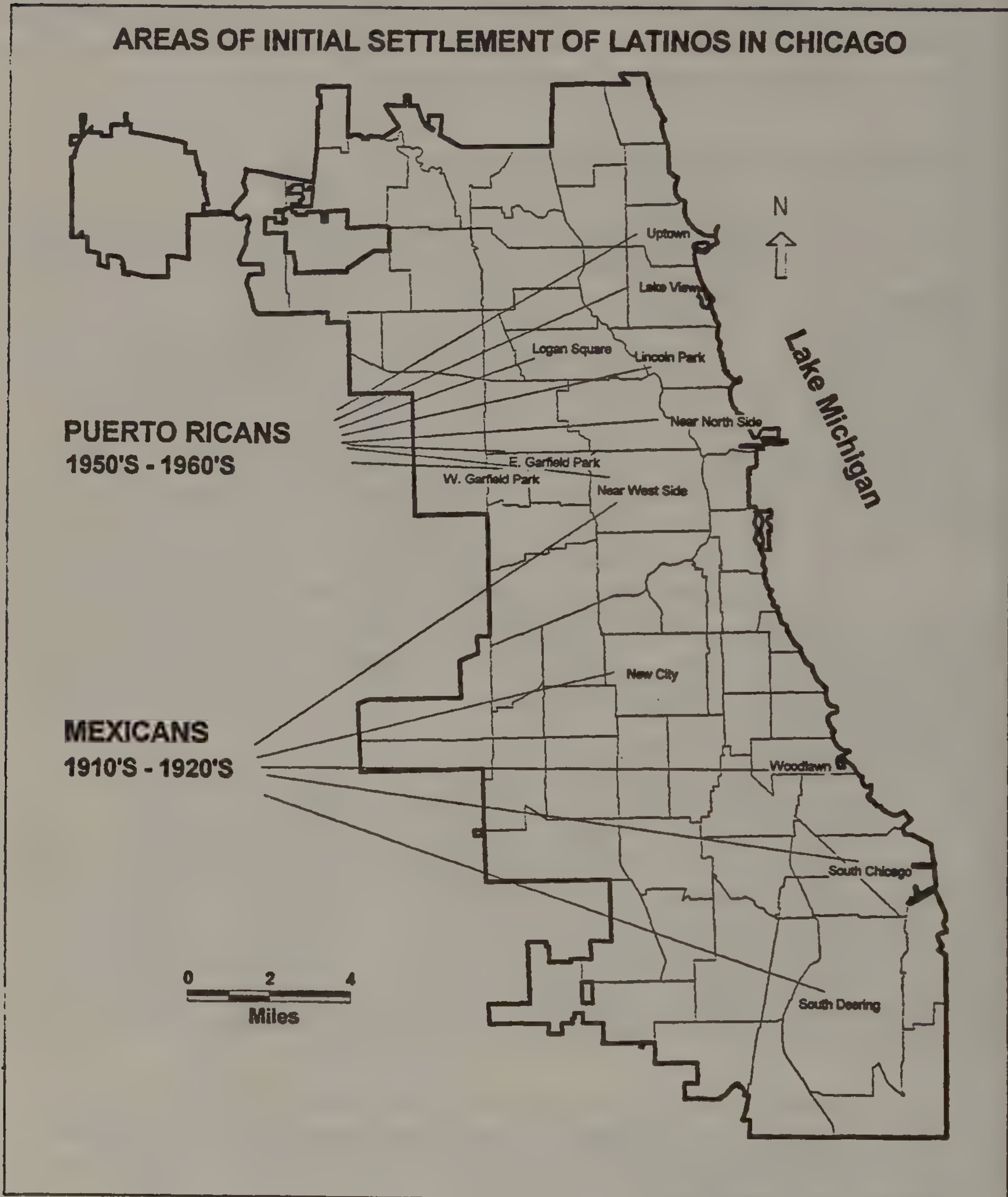
Taylor (1932) reports that while peer pressures forced landlords to exclude Mexicans from European ethnic — particularly resident homeowner — areas, profit interests motivated them to rent to Mexicans. Some dispersal to the fringes resulted from this. Meanwhile, opposition to Mexicans had a strong racial basis. Explanations about this rejection ranged from behavioral and cultural traits to color and national origin. Real-estate businesses claimed that Mexicans had a negative impact on property values. This is a clearly racist factor: as Mexicans move into an area, whites no longer want to live there. Property values depreciate as a result.[12] However, the industry managed to make up for this:

> Since the Mexicans have come in, the real estate values have declined to almost nothing. But the rental value of the buildings goes up $10 to $15 per flat; the Mexicans can only get in a very few places, and they have to pay what is asked (a realtor quoted by Redfield in Taylor 1932:222n)

Ethnic restrictions succeeded in clustering Mexicans in their own spaces (Taylor 1932). Latino origin facilitated this. The first Latino clusters were located in South Chicago-South Deering (next to the steel mills), the Hull House area (near the railroads and other central sources of employment), and in Back of the Yards (close to meatpacking houses; see Figure 1).

Available evidence suggests that clustering and speculation have been a constant for Latinos up to today. Analysis of the period 1930-70 points to the persistence and firm rooting of the initial Mexican clusters and the economic

FIGURE 1: Areas of Initial Settlement of Latinos in Chicago



position of Latinos vis-à-vis other groups (Año Nuevo Kerr 1976:12). Some dispersal was forced by fear of recognition and deportation in the 1930s; yet "continued residential restriction led to immobility" (95). A 1935 survey of persons on relief concluded that racial prejudice, segregation, labor segmentation, and cultural distinctiveness plagued the Mexican population in Chicago (Hughes 1937). As a result, through the 1930s, Mexicans "experienced greater economic difficulties than any other ethnic groups in the city, except Blacks" (Año Nuevo Kerr 1976:82).

Conditions did not improve in the 1940s. The 1948 Report of the Mexican Sub-Committee of the Metropolitan Welfare Council (Pax 1948) indicated that segregation and isolation in the three initial clusters, discrimination, and poverty prevented the advance of the group. Moreover, new immigrants continued to settle in or around the clusters (Año Nuevo Kerr 1976). Evidence also points to residential mixing among Mexicans of different social statuses, educational levels, and generations from the early 1920s to the 1970s (Año Nuevo Kerr 1976).

In the 1950s, highway construction displaced many from the eastern and northern borders of the Hull House area. In the 1960s, most of the area was cleared for a university campus, forcing residents to relocate. Displaced Mexicans joined a new cluster that had been forming out of the Hull House Mexican cluster,[13] to the south. More deteriorated than the previous one, this area — Lower West Side or Pilsen — soon became the first *majority* Mexican and Latino community in Chicago[14] and the basis for organization of the group. Latino population also increased steadily in the other clusters after the 1960s (Latino Institute 1983). Along with population increases came public neglect, slumlordism, and disinvestment. Negligible levels of Latino homeownership, overcrowding, arson, neglect, and high rents have characterized these areas in the last decades (Chicago Fact Book Consortium 1984).

Puerto Ricans also settled in small clusters according to their place of employment and the original housing provided by recruiters. They were first placed in cheap hotels, old passenger coaches, public housing, and similar housing. They then pushed their way into single-room-occupancy units and boardinghouses, shared quarters, and other run-down or vacant rental properties nearby. They gave a new lease on life to substandard housing. Initial clusters were located on the Near North Side and in Lincoln Park, Uptown, and Woodlawn (Martínez 1989; Padilla 1947; see Figure 1). Other groups settled later on the Near West Side; in West and East Garfield Park; and in Logan Square and along Division Street on the Near Northwest Side (Padilla 1987; see Figure 1). They were allowed in these areas because of vacancies generated by disinvestment and suburbanization. Blacks were also entering East and West Garfield Park and Woodlawn in large numbers at the time.

Like Mexicans, Puerto Ricans moved into rental housing in the most highly disinvested sections of those areas. Their stay in most of them was short. Highway construction and urban renewal displaced them in succession from the Near West and Near North sides and Lincoln Park (Jiménez 1974; Padilla 1987). Outnumbered by blacks, they also had to leave Woodlawn and East and West Garfield Park[15] (Padilla 1987). Puerto Ricans, then, consolidated along Division Street, in the West Town and Humboldt Park communities.[16] Unable to secure

white tenants, landlords reluctantly rented to them. Their conditions allowed very few to purchase. Owners, then, turned many units into rooming houses, kitchenettes, and other, often illegal, conversions for rent. Landlords speculated with the level of need and rejection of Puerto Ricans by limiting their choices through clustering and overcrowding in small, substandard, and unhealthy places with practically no maintenance or housing code compliance (Padilla 1987).

Discriminated against and displaced from their initial clusters, Puerto Ricans saw their move into the Division Street area as an opportunity to form a community they could call home away from home (Padilla 1987), one, indeed, from which they could not be so easily displaced. This became, then, the residence for Puerto Ricans of different occupations and economic and social status, with various generations living together. As in the case of Mexicans, Puerto Ricans faced housing discrimination from the beginning (Padilla 1987). Not only were they limited to clusters but "were forced to pay higher rents in those buildings that were open to them" (117). They suffered repeatedly from arson when landlords tried to recover their investment from insurance companies. One-third of the local housing was destroyed by abandonment, demolition, and fires between 1970 and 1978 (Padilla 1987).

Today, gentrification is forcing them out, into less accessible areas of recent disinvestment and "white flight" to the west and north.[17] Higher-income groups — whites mostly — have been incrementally replacing Latinos who can no longer afford the gentrifying neighborhood.[18] Meanwhile, Latinos are moving into adjacent areas of disinvestment that had become less desirable from their closeness to Latino concentrations and to white flight.

In short, the experience of Puerto Ricans has been one of clustering, displacement, and reconcentration. Also, Mexicans expanded out of their initial clusters or were displaced to reconcentrate nearby. Both were forced into this by the combined forces of real-estate speculation — illustrated here — and racism rather than as a result of upward mobility or integration. Mobility was largely horizontal. Forced and continuous moves also made many of them live in permanent transition. They have not displaced other ethnics. Latino-related speculation has been dictating the terms, rather than assimilation, choice, competition, and mobility. Latino solidarity has been maintained, indeed reinforced, by exclusion, segregation, and discrimination (Padilla 1987). In spite of some dispersal present since the beginning but largely restricted to the fringes of Latino concentrations (Taylor 1932 and our own research), movement out of these clusters has never had the massive character of European ethnic groups' dispersal.

From Arranged Immigration to Housing Discrimination and Segregation

The previous analysis has linked labor segmentation and residential clustering to the condition of being Latino. The present section summarizes other activities of the real-estate industry exemplifying this and further establishing its relationship with Latino settlement patterns. In particular, it agrees with the

claim that Latinos became a quasi-racial segment within a dual residential market system (Bradford 1991; Dennis & Pottinger 1988).

As early as 1933, in his ranking of groups by their impact on property values from positive to negative, Hoyt (1933) placed Mexicans last, behind blacks. The negative impact of these two groups on property values has always been assumed in real-estate theory and practice. This assumption has its roots in racism as the presence of these groups makes an area undesirable or less desirable for whites. It, in fact, has resulted in the rating of Latino and black areas as highly risky and in the disinvestment associated with such an assessment (Bradford 1991). It has facilitated group-based neighborhood change in Chicago and elsewhere as speculators use these groups to destabilize and turn around white areas (Boyer 1973; Bradford & Rubinowitz 1975; Bradford, Rubinowitz & McGowan 1975; Feins 1976). Finally, it has produced numerous racist schemes including steering, redlining, protective covenants, and segregation.

Before the 1949 Supreme Court decision that made restrictive housing covenants illegal, even public documents stressed separation and protective covenants as a condition for stability in land values (USFHA 1938, 1947). Practices of the industry did not change much after this decision (Berry 1979; Bradford 1991) or after the Fair Housing Act of 1968 (Kushner 1992). This was reflected in FHA maps circling minority areas to indicate that they were excluded from lending (Kushner 1992). Redlining of minority areas of residence in Chicago has been widely documented by the Woodstock Institute and others (Squires, DeWolfe & DeWolfe 1979). Using the data of the Institute, Shlay (1987, 1988) established that the higher the proportion of Latinos and blacks in an area, the lower the number of conventional loans. Steering of minorities to selected areas has been demonstrated in Chicago particularly through Fair Housing Audits (Austin Turner 1992; Galster 1992; Urban Institute 1991). Other mechanisms perpetuating housing discrimination and segregation have been identified in audits and other research in Chicago and elsewhere. They include withdrawal of conventional loans from Latino and black areas paired with the increase and abuse of FHA/VA-insured loans (Boyer 1973; Bradford, Rubinowitz & McGowan 1975; Feins 1976; Peterman & Sanshi 1991); speculative schemes associated with racial change (Boyer 1973; Bradford & Rubinowitz 1975); unequal outcomes in the search for housing compared to whites with the same incomes (Galster 1992); unequal treatment in housing transactions (Austin Turner 1992); failure to serve minorities adequately (Weink 1992); discrimination in terms and conditions for rental housing, sales and marketing of property, and financial assistance for home buyers (Austin Turner 1992); minority purchases with cash or on contract (Bradford 1991); poor law enforcement (Dedman 1988; Kushner 1992); and discrimination in housing availability (Downs 1992).

Wachter and Megboulegbe (1992) relate discrimination against Latinos and blacks in housing to discrimination in labor markets via income. Similarly, Alba and Logan (1993) have established that even in the suburbs blacks and Latinos, in that order, exhibit a great deal of separation from whites. They also concluded that whites with distinguishing characteristics are much more likely to live together than with blacks or Latinos with similar nonethnic characteristics. Studies of segregation have repeatedly demonstrated that blacks and Latinos

have the highest rates of segregation (Massey & Denton 1987). While less segregated than blacks (White 1987), Latinos have consistently undergone extreme segregation in Chicago (Orfield & Gaebler 1991).

Two other factors suggest that the above trends will carry into the future: the first is a behavior-and-expectations mechanism and the second is the increasing subtlety of discriminatory practices. A long history of discrimination and segregation has produced expectations and beliefs that keep groups in their place. Fears are internalized and turned into behavior that perpetuates the status quo. Practices become then "constraints" in the settlement of groups. Second, discrimination and segregation take place today under an appearance of neutrality that the industry has been perfecting over the years. Control of listings, targeted marketing, claims of affordability, minimum loan amounts, higher closing points for smaller loans or risky areas, restrictions on financing of older houses or homes in racially mixed areas, informal covenants, poor enforcement of regulations, restrictions on children, claims that an apartment was just rented or that a house is under contract are good examples of practices often used to disguise discrimination (Austin Turner 1992; Berry 1979; Bradford 1991; Luttrell 1970).

In short, the condition of Latinos has been exploited and perpetuated by the real-estate industry in Chicago and elsewhere. It has been used to destabilize white areas or produce so-called white flight, to extract exorbitant profits from residential change, to create artificial pockets of scarcity by limiting the areas and types of housing available to Latinos, to devalorize areas,[19] and to maintain segregation (Squires et al. 1987; Suttles 1972). These factors, in turn, lead to separation, concentration in areas of disinvestment and white flight, and clustering and reclustering but, very particularly, high levels of manipulation in the residential market.

Recent Settlement Trends and Housing Conditions of Latinos in Chicago

This section completes the story of Latino concentration and segregation in Chicago with the introduction of relevant statistics and other current evidence. It also argues that initial, supposedly temporary situations have become permanent for Latinos in the city. Similarly, it shows that, overall, their housing situation is very similar to that of blacks in the city. This analysis draws on statistics from the U.S. Census of Population and information from other studies. It profits from the author's knowledge of Latino areas in Chicago, from formal and informal field observations over the last fifteen years, and from multiple conversations with persons in the real-estate industry, Latinos, and others.

With the exception of South Chicago, the traditional areas of Mexican and Puerto Rican concentration have increased their Latino population in recent decades (see Table 1).[20] Additional Latino settlement has taken place in the original clusters or has spilled over into their surrounding areas (Figure 2). Latino population increases have often been accompanied by increases of blacks; these, in turn, have led to decreases of Latinos next.[21] The process of disinvestment and white flight slowly opened the West Town, Humboldt Park, and

TABLE 1: Communities of Major Latino Concentration — 1990 Central Indicators[a]

Community	Percent Latino 1980	Percent Latino 1990	Percent below Poverty	Latino Percent of Total below Poverty	Per Capita Income (All)	Per Capita Income, Latinos
Uptown	23	23	31.3	27.2	12,389	6,374
Lincoln Square	11	23	13.2	38.0	13,091	8,235
North Center	19	27	12.2	37.9	14,132	8,505
Albany Park	20	32	17.5	40.7	10,273	7,418
Irving Park	9	24	10.2	29.0	13,721	10,256
Belmont-Cragin	6	30	8.4	40.4	11,938	9,032
Hermosa	31	69	17.4	82.1	9,287	7,880
Avondale	21	38	17.4	44.9	11,228	7,741
Logan Square	52	66	26.4	71.7	8,846	6,584
Humboldt Park	41	49	33.8	41.9	6,905	6,099
West Town	57	62	31.9	71.0	8,258	5,595
South Lawndale	74	85	24.0	86.0	6,307	6,138
Lower West Side	77	88	27.8	91.3	6,467	5,680
New City	36	39	34.1	31.7	6,607	6,118
Bridgeport	21	26	14.6	27.4	10,715	7,891
McKinley Park	16	40	10.9	44.5	10,682	8,075
Brighton Park	15	37	12.6	45.0	11,000	8,076
Gage Park	11	39	11.4	40.7	11,207	9,326
Chicago Lawn	11	28	18.3	24.7	10,144	8,217
South Chicago	39	33	25.1	38.3	8,885	6,609
South Deering	24	28	18.4	40.2	10,038	7,492
East Side	13	40	13.0	55.0	10,716	7,830
Rogers Park	12	20	19.5	23.3	12,603	7,317
Total	na	43	21.7	50.9	na	na
City of Chicago	14.1	19.6	21.6	21.7	12,899	7,438

[a] Source: City of Chicago Department of Planning and Development. *Social and Economic Characteristics of Chicago's Population,* 1992.

Logan Square areas to blacks.[22] With the exception of the Hull House area, Latino concentrations, thus, expand today out of the original cores established in the 1910s and 1920s by Mexicans and in the 1960s and 1970s by Puerto Ricans (Figure 2 and Table 1). In Pilsen, Mexicans moved southwest into South Lawndale[23] to reach 85% of the area in 1990. They expanded further into the adjacent suburbs of Cicero and Berwyn to the south and west. Latinos have been moving east of South Chicago into the East Side[24] and into the suburbs. Blacks have followed Latinos into the East Side. While no blacks were registered on the East Side by the 1990 census, an increasing number of them has been entering since.

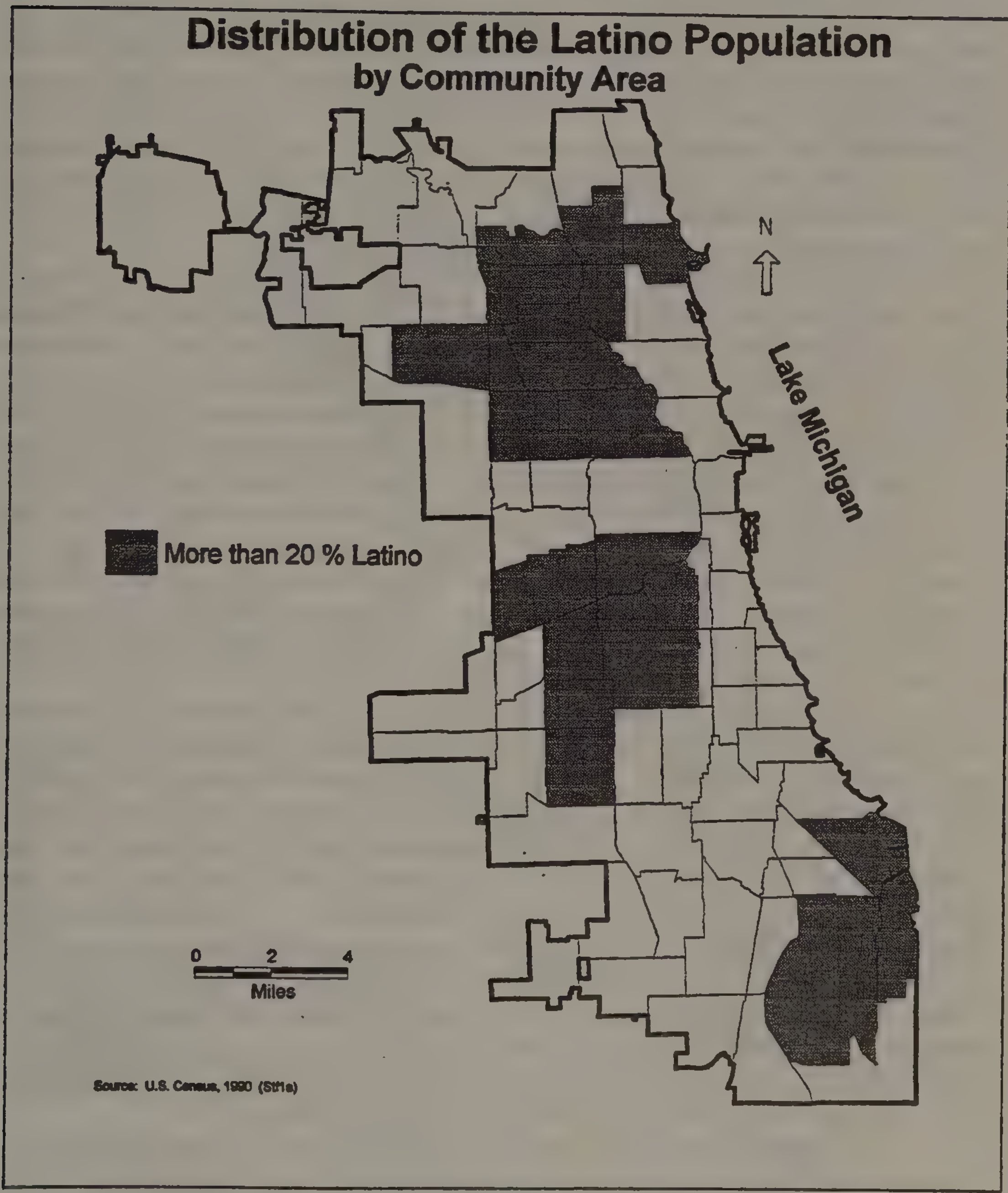
Similarly, Latinos expanded southwest from Back of the Yards into neighboring areas such as Brighton Park, McKinley Park, Bridgeport (with 0% blacks), and Gage Park[25] (1% black in 1980 and 5% in 1990). Not only are Latinos increasing their share in these white areas, but they are forming a buffer zone between whites to the west and blacks to the east. The Division Street core in the near northwest has also been expanding into adjacent areas to the west and to the north including Hermosa, Avondale, Albany Park, North Center, Lincoln Square, Belmont-Cragin, and Irving Park.[26] Expansion has reached into nearby old suburbs such as Addison and Melrose Park. These communities had no or less than 3% blacks in 1990. West Town, Hermosa, and Belmont-Cragin separate the black West Side from white and mixed — white, Latino, and Asian — areas to the north and west. Gentrification in Uptown, Lakeview, and West Town has also pushed many Latinos into these areas.

Other Latinos have been entering Mexican and Puerto Rican areas since the late 1950s. These two groups also lived with each other in South Chicago (with the recruitment of Puerto Ricans for work in the steel mills after World War II) and in West Town (where Mexicans started moving in such numbers since the 1960s that, by 1990, they had surpassed Puerto Ricans). Even though areas such as the near southwest and sections of the near northwest are identified as Mexican and Puerto Rican respectively, Latino neighborhoods are becoming increasingly mixed, sharing experiences and problems with nationals from other Latin American countries.

As mentioned before, some areas have been losing Latinos to gentrification[27] since the 1970s, but particularly in the 1980s and 1990s. Such is the case of the east end of Pilsen, which has gained whites while losing Latinos. Latino clusters are near extinction in gentrified Lakeview — adjacent to Lincoln Park — particularly in the middle of the community. In West Town, gentrification has been advancing steadily in Wicker Park, Bucktown, and the southeast sections with substantial reductions of Latinos. Similar displacements have been taking place in Logan Square and Uptown.

Currently, 83% of Latinos are concentrated in three main contiguous areas, namely the near to mid northwest, the near to mid southwest, and the far southeast (see Figure 2). The northwest area includes 11 communities with populations that are 20% or more Latino — Latinos are 20% of the population of Chicago — which comprise 45.1% of the city's Latino population. The southwest area includes 8 communities and comprises 33% of the Latinos in Chicago. And the far southeast includes 3 communities with 5% of the city's Latinos. Added to these is an area in the far north end, Rogers Park, with 2.2%,

FIGURE 2: Distribution of the Latino Population



for a total of 85% of Latinos concentrated in 23 communities out of the 77 communities that comprise the city (Table 1).

The areas of concentration of Latinos and, more specifically, the subareas of Latino clustering are among the most deteriorated and oldest in Chicago. Out of 41,667 housing units lost by the city to deterioration between 1980 and 1983, 45% are located in the 23 communities of highest Latino concentration, particularly in their Latino and black subsections (Puente & Vásquez 1989). Considered ports of entry or lower-income communities, all of them have lower mean household incomes than the city and all but three have per capita incomes below those of Chicago. In each of them, Latinos have lower per capita incomes and more persons per room than the total population.

Latinos are not integrated with other groups in these areas. They tend to live in subclusters within census tracts, particularly in rental sections. They are less segregated than blacks in the sense that they live interspersed with other groups, particularly in areas in transition. Yet their clustering in specific blocks, buildings, or subsections of communities is visible to anyone. This is reflected in Figure 3, which shows a core of high concentration in or near their original clusters surrounded by contiguous census tracts with less concentration.

Ten percent more non-Latino households live in owner-occupied units in these areas than Latino households (Table 2). In other words, not only do Latinos live in the worst housing, but in rental housing. Notice that these are low-income areas. Hence, Latinos continue having the worst conditions even in the low-income areas in which they are concentrated in Chicago. This may also reflect the transitional nature of many such areas.

An examination by census tract shows that in 1990 Latinos paid more rent than non-Latinos in 360 of the 764 census tracts in the city. The average difference was $69 or 26% more than the rest. In contrast, they paid less rent in 404 tracts. The difference here was lower, at an average of $43 or 15%, suggesting that those who paid less were closer to the average than those who paid more. Nearly half of the tracts (49.4%) in which Latinos paid more rent were located in areas with 20% or more Latinos compared to 34% of those in which they paid less. Higher rents in areas of clustering may be associated with local scarcity resulting from concentration or from adjustments to Latino-related losses in housing prices. Lower rents in areas of lesser Latino presence may be related to the lower condition of housing rented to Latinos. As suggested by a study of the *Chicago Reporter* ("Hispanics and Asians" 1990), Latino and Asian areas are also the hardest hit in the city by rent increases.

A few citywide indicators help complete this picture. The percentage of owner-occupied housing units in 1983 was 31.4 for Latinos, compared to 33.8 for blacks and 48.8% for whites. A total of 90% of Latinos, 65% of blacks, and 61% of all households lived in housing built before 1939 (Puente & Vasquez 1989). Eighteen percent of Latinos lived in overcrowded housing in 1983 compared to only 10% of blacks and 7% of the total population (Puente & Vásquez 1989). While 58% of Latino households moved between 1980 and 1983, the figure for both blacks and all households was only 38%. A remarkable 71% of Latino renters moved in the period, compared to 50% of blacks and 55% of all renters (Puente & Vásquez 1989). More than 42% of Latino renters qualified for public housing in 1983 — compared to 33% of blacks and 23% of all renters. Only 1.7%

FIGURE 3: City of Chicago by Race and Ethnicity

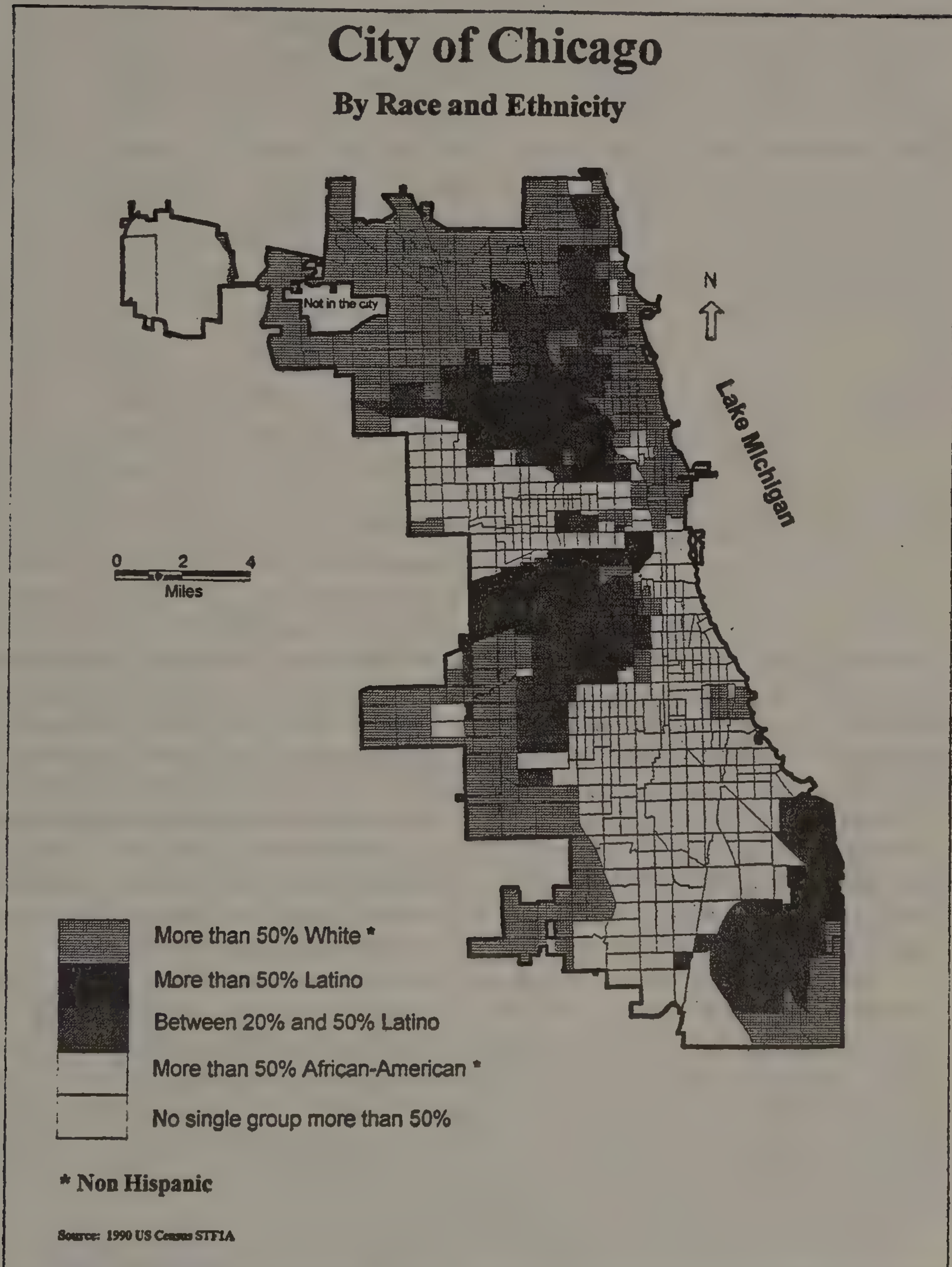


TABLE 2: Percentage of Owner-Occupied and Renter-Occupied Units

Area	Occupancy	Latino	Non-Latino	Total
Near Northwest	Owner	26.2	43.2	38.4
	Renter	73.8	56.8	61.6
Near Southwest	Owner	36.2	na	44.1
	Renter	63.8	na	55.9
Far South	Owner	46.8	61.8	57.9
	Renter	53.2	38.2	42.1
Total, Three Areas	Owner	31.1	46.2	41.4
	Renter	68.9	53.8	58.6
Chicago	Owner	31.4	43	41.5
	Renter	68.6	57	58.5

[a] Source: U.S. Census of Population, 1989, STF3 Files.

of all public housing units were occupied by Latinos, 85.9% by blacks, and 11.9% by whites (Puente & Vásquez 1989). In that same year, only 2% of Latino renters lived in subsidized housing in the city compared to 20% of all blacks and 10% of all renters. In the 10 communities with the highest Latino percentages in 1980 and 60% of the Latino population of the city, only two out of ten households were occupied by Latino homeowners (Puente & Vásquez 1989).

Hence, Latinos were concentrated in the oldest areas and housing stock in the city, had the lowest homeownership rates, the highest concentration in mid-size, multifamily rental buildings, the lowest value homes, and the highest levels of overcrowding and residential mobility among all groups. They also paid a higher share of their income in rents and mortgages than whites.

This analysis and these findings are from within the city limits of Chicago, and we need also to examine the details from the suburbs. From an aggregate analysis, Massey and Denton (1988) argue that the levels of segregation in the suburbs are lower for all groups; while the segregation of blacks is still high, that of Latinos and Asians ranges from low to moderate. Examining this situation in detail for the suburbs of Chicago is beyond the scope of this article. Our preliminary assessment, however, suggests the possibility that many of the situations described for Chicago hold there as well.

First, large suburban Latino concentrations are often expansions of clusters established around World Wars I and II. Their process of formation is similar to the one described for Chicago. In fact, they resulted from the direct importation of workers by railroad and manufacturing employers (Año Nuevo Kerr 1976). Second, Latino increases have taken place particularly in old suburbs of white flight, disinvestment, and deindustrialization. Even though suburban Latinos are slightly better off than Latinos in the city, their suburbanization is proceeding

along the lines of clustering in the most deteriorated rental areas (Latino Institute 1993). Third, Latino concentrations may not be easily captured by analyses at the census-tract level because they are often located in smaller enclaves around multifamily rental housing or because their suburban growth is too recent to show definite patterns. Many of their enclaves have become important reservoirs of low-wage labor for the suburban economy. Preliminary research in the suburbs of the northwest points to the clustering of Latinos there (Tobias & Roy 1993). Our own analysis shows, in nine of the ten suburbs with the largest Latino populations, concentrations in a few census tracts far above the overall Latino percentage. In eight of them (Addison, Aurora, Blue Island, Chicago Heights, Cicero, Joliet, Melrose Park, and Waukegan), with Latino populations ranging from 13% to 37%, between 39% and 74% of Latinos are concentrated in three, mostly contiguous, census tracts. Then, in the suburb of West Chicago, 30% Latino, 61% of the Latino population is clustered also in one single tract.

Fourth, some of the most dramatic growth has taken place in suburbs immediately contiguous to the city. In fact, they continue the expansion of current Latino concentrations in Chicago. Fifth, incidents of discrimination seem to be growing with increases in the population and concentration of Latinos. The cases of Cicero, Hodgkins, Waukegan, and Addison where Latino groups are challenging efforts by these municipalities to displace them through renewal and code enforcement in their areas of residence illustrate this.[28] A study by the *Chicago Reporter* ("Suburban Housing" 1995) of six suburbs where the Latino population doubled in the 1980s concluded that Latinos were getting a disproportionate share of code enforcement efforts in these suburbs. Sixth, many suburbs with significant Latino populations have large and growing black concentrations, often bordering Latino clusters. Seventh, the trend in some suburbs is toward a Latino majority in the near future. In short, this evidence calls for analyses of the suburbs at a disaggregated and qualitative level to identify elements of segregation and discrimination that cannot be captured by the indexes and other aggregate analyses used in support of the ecological model.

Conclusion

This analysis has identified many elements of discrimination in the Latino residential experience in Chicago. It suggested the existence of a "Latino rent" and a high level of manipulation of their settlement process. It illustrated the extreme vulnerability and instability of Latino communities associated with their status in the U.S. and the resulting "racialization" of their settlement. It argued that Latino status is a source of segregation and speculation in the city's housing market. As Latino settlement intensifies or expands, adjacent white areas become less desirable and are progressively disinvested and thus opened up to Latinos. Once Latinos enter them in large numbers, they are also opened to blacks and the process of transition and disinvestment intensifies. Along the way, Latinos often become a buffer between blacks and whites. They provide owners with the opportunity to recover their investment, in fact, to engage in

discriminatory speculation based on the Latino condition. From this perspective, it could be argued that Latinos end up facilitating, if not financing, white flight. On these bases, the article questions Massey's claim that Latinos are "ethnics" properly represented by the ecological model.

Our research also suggests that clustering has been strongly associated with Latino status in Chicago. While choice and common characteristics are important factors, exclusion associated with a Latino status has intensified and perpetuated it. Similarly, the article suggests that dispersal is not followed by assimilation as in the ecological model but by reconcentration. It is often the result of displacement, disinvestment, overcrowding, and the instability associated with low ownership, immigration status, back and forth migration, and unstable unemployment related to Latino status. Residential mobility, then, is largely horizontal. Higher-income Latinos often move to the fringes of Latino areas while maintaining a close association with their clusters of origin. Like blacks, others may be able to move into white areas individually. As their numbers increase, however, their presence is increasingly resented and may set up the process of white flight.

This article linked the condition of Latinos to the terms of their immigration and the domination associated with their origin and minority status. While, generally, European ethnics had high levels of mobility within and between industries and occupations leading to socioeconomic mobility, Latinos were segmented into marginal, unskilled jobs that froze them at the bottom. In the model, residential and socioeconomic mobility are closely related. In contrast, subjugation and "racialization" have denied Latinos substantial economic assimilation and have subjected them to multiple forms of segregation unaccounted for in the ecological model.

Our research points to the importance of qualitative analyses for identification of explanations and elements that statistical tests often miss. It, in fact, suggests interpretations of Latino settlement different from those of the ecological model. While the levels of Latino segregation are lower than those of blacks (Massey & Denton 1987, 1993), this may be the result of the higher mobility and instability discussed rather than of the integration and mobility implied by the model. "Integration" may be a phenomenon of an area's transition. The high resilience of original Latino clusters may be related to anti-Latino prejudice and exclusion. Dispersal into adjacent white areas is often only the beginning of the turnover process. Higher-income Latinos may be the ones moving first, as they have the choice of homeownership and provide whites with the opportunity to exit areas adjacent to Latino clusters. As this movement to the fringes intensifies, more whites leave and rent to Latinos. In fact, a high correlation between socioeconomic status and mobility away from Latino cores may result from the pioneer role of higher-income Latinos (Aguirre, Schwirian & LaGreca 1980; Jackson 1981; Massey 1979; Valdez & Jones 1985).

Thus, the settlement of Latinos in Chicago differs in significant ways from the pattern and assumptions of the ecology model. Unlike Europeans, Latinos bear a permanent minority condition that has translated into exclusion, economic immobility, and manipulation. This is reflected in a settlement process deeply colored by discrimination. U.S. society has extended racism to Latinos; their settlement reflects this. Moreover, ongoing national economic restructuring

pointing to a shrinking of middle classes and opportunities, to a highly polarized labor market, and to the intensification of the disadvantages of minorities (Harrison & Bluestone 1988) casts further doubt on the future applicability of the ecological model to Latinos.

Meanwhile, the dynamics of racism have separated Latinos from other minorities. Once excluded, Latinos and blacks have turned their segregated communities into a basis for struggle, often to the exclusion of each other. As a result, the pattern of separation is intensified. As disinvestment destroys black areas, blacks start pushing outside their borders and entering areas frequently opened first to Latinos. It is not as much a matter of choice and succession as a matter of segregation and market manipulation. We should not equate these dynamics to "ethnic" succession and upward mobility.

The continuous influx of Latinos may feed into their areas of first settlement and reinforce their concentration. This factor needs to be examined in detail. Evidence suggests that in nearly a century of Latino immigration to Chicago, patterns of segregation and dispersal have been continuous. Residential assimilation has never been close to the levels of European immigrants. Tension and real-estate speculation related to Latino status may be fed by new entrants. Low levels of tension associated with recent or continued European influx, though, point to prejudice in the case of Latinos. This prejudice, indeed, extends to middle-class and U.S.-born Latinos and is reflected in the reaction to increases in the proportion of Latinos in white areas. Suburbanization of Latinos may conceivably fit the model. Limited evidence available for the Chicago suburbs, however, suggests that the suburbs are experiencing the same general trends of clustering, segregation, and dispersion as the central city. Detailed testing of these claims is necessary to gain a complete understanding of the settlement experience of Latinos in Chicago. Finally, other models need to be developed or further examined for alternative explanations of the Latino experience.

Notes

1. Discrimination against Latinos has ranged from subtle forms of exclusion to more open ones of labor segmentation (Meléndez 1988; Meléndez, Rodríguez & Figueroa 1991) and immigration status (Kirstein 1977; Kiser & Kiser 1979). While some of them have been documented, most others have not. In fact, research on Latinos in the U.S. lags far behind that of blacks and European immigrants.

2. Little attention has been paid to the relationship between U.S. domination of Latin America, the terms of migration, and the conditions of Latinos in the U.S. Latin Americanists are aware of the use of migration as a tool in the domination of the region. Analysts of the Latino experience in the U.S. know about the coalition of the U.S. government and employers for importation of Latinos on terms leading to their labor segmentation and overexploitation. The combination of their actions has produced a highly vulnerable Latino population with conditions, status, and treatment similar to those of American Indians and blacks.

3. We focus on Puerto Ricans and Mexicans because they are 65 and 23% respectively of the Latino population in Chicago. We are aware, however, that the individual cases of other Latino subgroups need to be examined in detail, as their experiences may differ from the experience identified here.

4. These were nonunion, labor-intensive industries relying heavily on immigrant labor. Latinos not only were assigned unskilled jobs but had minimal mobility through the ranks (Rosales &

Simon 1978; Taylor 1932). This translated into a high concentration of Latinos in such industries and occupations — as late as the 1980s (Año Nuevo Kerr 1976; Betancur, Córdova & Torres 1993).

5. Employers encourage Latino employees to bring friends, relatives, or others they know. This gives them some guarantee that the recruits are of the same quality. Moreover, the recommending party makes sure that the other person works out.

6. One case in point is the Bracero Agreement between the U.S.A. and Mexico in World War II to import Mexicans for temporary work in occupations and industries for which there was a labor shortage. The agreement explicitly indicated that braceros were not to compete for jobs for which U.S. labor was available (Kirstein 1977:15-24). In this way, Mexicans were once again limited to low-end jobs in labor-intensive industries. Labor segmentation ensued as employers continued pursuing Mexicans for these types of jobs.

7. Mexicans were blamed in the Great Depression for taking "the jobs of Americans" (Rosales & Simon 1978). "Repatriation" was largely a result of this. Undocumented workers have been blamed and selectively raided and deported in every economic downturn. Proposition 187 in California and the wave of sentiment against the undocumented and against immigrants in general is the most recent expression of this. In the 1940s, on the occasion of a confrontation between servicemen partying in a Mexican neighborhood in Los Angeles and local youth, the national press blamed Mexicans and branded them as subversive, unredeemable, and opposed to the war. Special outreach efforts took place next in U.S. cities, under the assumption that Latinos, especially youth, were particularly vulnerable to subversive influences (Año Nuevo Kerr 1976). Latinos have been stereotyped as violent, noncompetitive, lacking the proper values and drive for success, etc. And their condition has been often blamed on this.

8. Acquired in the Spanish-American War of 1898, Puerto Rico became a heaven for U.S. investment and a source of cheap labor for the U.S. economy. The U.S. extended the model of dependent development of Latin America to the island. Strong independence sentiments developed during Spanish tenure of the island were then directed toward the U.S. The conflict with the U.S. and the model of development became a continuous source of tension in the island.

9. In its Statement of Motives of 1947, Puerto Rico announced that it would cooperate with mainland agencies in the settlement of immigrants and guide them to areas of labor demand (Maldonado 1987). An agency of the Department of Labor of Puerto Rico was established in Chicago in 1949 to facilitate the recruitment of Puerto Ricans for work in the city and the Midwest (Martínez 1989). Private recruiters such as Castle, Barton and Associates and the Friedman Farm Labor Agency imported Puerto Ricans for firms such as the Chicago Hardware Foundry Company and U.S. Steel. Castle also provided households with maids from the island (Martínez 1989).

10. Arranged migration refers to the unique terms of Latino importation discussed in this section. Arrangements such as temporary admissions or undocumented migration produced a highly disadvantaged group of "illegal," second-class, or transient workers. Importation for strikebreaking set them against white workers while recruitment for marginal industries and occupations turned them into a highly segmented labor pool. U.S. actions in Latin America generated political and economic refugees that were often used to wage counterrevolutionary drives in the region. Latino migration often became a revolving door, and Latinos became immigrants of last resort with highly unstable or vulnerable statuses, "illegal aliens," second-class immigrants, or an imposed presence on America that was resented by other workers and groups.

11. Evidence indicates that the importing companies had to play a role in housing Mexican workers in order to guarantee their migration. Housing scarcity particularly near the areas of employment was a factor. In the case of workers recruited to replace strikers, housing them inside the company grounds was crucial.

12. The words of a realtor quoted by Taylor (1932) illustrate this: "Mexicans depreciate property values and take the jobs of others. When the Mexicans come in, the others say they want to sail out" (222).

13. Interviews of Mexican residents of the Hull House area reveal that urban renewal and real-estate agents had steered them to Pilsen.

Mexicans stabilized and started reversing the decline. Today, it is among the oldest and most overcrowded areas of the city. Slightly over half of its units are overcrowded, compared to 27.3% for Chicago as a whole. Three-fourths are rental units. Even though Latinos were nearly four-fifths of residents in 1980, they owned under half the owner-occupied units.

15. As blacks gained control, they "marked these communities off" as their "exclusive domain" (Padilla 1987:85). Puerto Ricans felt that "Blacks would never voluntarily yield any of the benefits or resources" of such places (85). "Racial problems or social misunderstandings with Blacks resulted in the movement of Puerto Ricans from these areas" (86).

16. Again, Puerto Ricans had lived on the Near North Side, close to downtown. With urban renewal, they were pushed west into the adjacent communities of West Town, Humboldt Park, and Logan Square. As the South Side and East Side areas they originally lived in became black, in the 1970s they joined the cluster on the Near Northwest Side in large numbers to form their own neighborhood.

17. Promoters of gentrification in Wicker Park, a subsection of West Town, explain that they acted to save the architectural and historical value of the area from the deterioration associated with Puerto Rican residents.

18. Our fieldwork indicates that former homeowners were able to buy in adjacent areas because of the "high prices" obtained from their properties. Often, however, they purchased smaller houses. Others had to double or triple up with relatives to buy. Tenants reported moving many times near their original quarters, often having to double up because of the higher rental price of the new units. The result of this has been increased overcrowding in subsections of adjacent communities.

19. As Latinos move in, the value of housing decreases along with the desirability of the area. Disinvestment intensifies or follows, further devalorizing the area. Devalorization may then create an opportunity in the gap between the rent currently realized and the one that could be realized through reinvestment (Smith 1979).

20. In Pilsen, Latinos went from 55% of the residents in 1970 to 88% in 1990; in Back of the Yards from 13% to 39%; in South Chicago from 26.1% to 33%; in West Town from 39.1% to 62%; in Logan Square from 17.8% to 66%; and in Humboldt Park from 15.5% to 49%.

21. In Back of the Yards, blacks surpassed Latinos in the period 1980-90, growing from 22 to 41%. In South Deering, they went from 54% to 59% (Latinos increased by a mere 241 persons here). Blacks in South Chicago went from 22.4% in 1970 to 48.8% in 1980, and to 61% in 1990, surpassing the declining Latino population. South Chicago lost 5,656 people in the 1980s.

22. Blacks, in fact, went from 35% to 49% in Humboldt Park between 1980 and 1990 to equal the Latino population in 1990. They remained stable at 9% in West Town and went from 2% to 5% in Logan Square. This expansion is related to the increasing deterioration of black communities to the south — the West Side of Chicago.

23. South Lawndale has one of the oldest housing stocks in Chicago. Having achieved majority in Pilsen, Latinos moved southwest into South Lawndale to become 31.9% in 1970, 74% in 1980, and 85.2% in 1990. South Lawndale is also predominantly a rental community with only 34% of its units owner occupied, over half of them (57%) owned by Latinos in 1980. Pilsen and South Lawndale have among the lowest vacancy rates and the highest rates of overcrowding in the city.

24. Traditionally a stable white working class area in which "homes are handed down from generation to generation" (Chicago Fact Book Consortium 1984:138), the East Side has become increasingly Latino (from 13% in 1980 to 40% in 1990).

25. Latinos are highly concentrated in subsections of these areas, forming a continuum with other Latino areas or with subsections of adjacent clusters. Typical resident-owner areas, Brighton Park, McKinley Park, Bridgeport, and Gage Park have been increasing their rental housing as whites move out. Traditionally closed to nonwhites and the scenario of many racial conflicts, the last two areas have opened up as the combined result of white decline, disinvestment in the city and the growth of the Latino population.

26. Hermosa, Avondale, Albany, North Center, Lincoln Square, Belmont-Cragin, and Irving Park are all adjacent, traditionally white communities. As Uptown to the east and Logan Square and Humboldt Park to the south increased their Latino and black populations, whites started leaving and opening the door to others. With the exception of Belmont-Cragin, all these

areas are largely rental. Albany and North Center have an extensive stock of multifamily buildings. The others have a mix with a predominance of small two- to four-unit buildings. Latinos are concentrated in rental units, usually in the most deteriorated sections.

27. Gentrification in Chicago has taken place mostly in Latino areas. Location near downtown and the architectural value of their housing made them particularly attractive. The vulnerability and high levels of Latinos renting facilitated this.

28. As was noted in the 7 October, 1994 edition of the *Chicago Tribune,* Latinos filed a lawsuit in Addison claiming that redevelopment plans for a section of the municipality are discriminatory and violate the civil rights of Latinos. This suit was joined by the U.S. Justice Department, which accused Addison of violating the federal Fair Housing Act. Meanwhile, in 1993 this same department sued the suburb of Cicero on grounds that its occupancy ordinance was discriminatory against Latinos, as was described in the *Chicago Reporter*, September 1995.

References

Acuña, Rodolfo. 1988. *Occupied America: A History of Chicanos.* 3d ed. Harper & Row.

Aguirre, Benigno E., Kent P. Schwirian, and Anthony J. LaGreca. 1980. "The Residential Patterning of Latin American and Other Ethnic Populations in Metropolitan Miami." *Latin American Research Review* 15:35-63.

Alba, Richard D., and John R. Logan. 1993. "Minority Proximity to Whites in Suburbs: An Individual-Level Analysis of Segregation." *American Journal of Sociology* 98:1388-1427.

Año Nuevo Kerr, Louise. 1976. "The Chicano Experience in Chicago, 1920-70." Ph.D. diss., University of Illinois at Chicago.

Austin Turner, Margery. 1992. "Discrimination in Urban Housing Markets: Lessons from Fair Housing Audits." *Housing Policy Debate* 3(2):185-215.

Bach, Robert L. 1978. "Mexican Immigration and the American State." *International Migration Review* 12:536-58.

Barrera, Mario. 1979. *Race and Class in the Southwest: A Theory of Racial Inequality.* University of Notre Dame Press.

Bean, Frank D., and Marta Tienda. 1987. *The Hispanic Population of the United States.* Russell Sage Foundation.

Berry, Brian J.L. 1979. *The Open Housing Question: Race and Housing in Chicago, 1966-76.* Ballinger.

Berry, Brian J.L., and Frank Horton. 1970. *Geographic Perspectives on Urban Systems.* Prentice-Hall.

Berry, Brian J.L., and John D. Kasarda. 1977. *Contemporary Urban Ecology.* Macmillan.

Betancur, John J., Teresa Córdova, and María Torres. 1993. "Economic Restructuring and the Process of Incorporation of Latinos into the Chicago Economy." Pp. 109-32 in *Latinos in a Changing U.S. Economy,* edited by Frank Bonilla and Rebecca Morales. Sage.

Blauner, Robert. 1969. "Internal Colonialism and Ghetto Revolt." *Social Problems* 16:393-408.

Bonilla, Frank, and R. Campos. 1981. "A Wealth of Poor: Puerto Ricans in the New Economic Order." *Daedulus* 110:133-76.

Boyer, Brian D. 1973. *Cities Destroyed for Cash: The FHA Scandal at HUD.* Follett.

Bradford, Calvin. 1991. "Never Call Retreat: The Fight against Lending Discrimination." Pp. 5-18 in *Credit by Color: Mortgage Market Discrimination in Chicagoland.* The Chicago Area Fair Housing Alliance.

Bradford, Calvin, and Leonard Rubinowitz. 1975. "The Urban-Suburban Investment-Disinvestment Process. Consequences for Older Neighborhoods." *Annals of the American Academy of Political and Social Science* 422:77-86.

Bradford, Calvin, Leonard Rubinowitz, and James McGowan. 1975. "Sample of HUD-FHA Single Family Home Commitments in the Chicago Metropolitan Area." Pp. 383-96 in *The Role of Mortgage Lending Practices in Older Urban Neighborhoods,* edited by the Urban-Suburban Investment Study Group. Northwestern University Center for Urban Affairs.

Bullock, Charles S., III, and Susan A. Manus. 1990. "Structural Features of Municipalities and the Incidence of Hispanic Councilmembers." *Social Science Quarterly* 71:665-81.

Centro de Estudios Puertorriqueños. 1979. *Labor Migration under Capitalism: The Puerto Rican Experience.* Monthly Review Press.

Chicago Fact Book Consortium. 1984. *Local Community Fact Book. Chicago Metropolitan Area, 1980.* Chicago Review Press.

Chicago Reporter. 1995. "Suburban Housing Inspectors Crack Down on Latinos." September.

____. 1990. "Hispanics and Asians Pay More Rent for Community." June: 3-5.

Corwin, A.F. 1978. *Immigrants . . . and Immigrants: Perspectives on Mexican Labor Migration to the U.S.* Greenwood Press.

Dedman, William. 1988. "The Color of Money." *Atlanta Journal-Constitution,* May:1-16.

Dennis, Warren L., and J. Stanley Pottinger. 1988. *1980 Federal Regulations and Banking: Redlining and Community Reinvestment Analysis.* Warren, Gorham & Lamont.

Downs, Anthony. 1992. "Policy Directions Concerning Racial Discrimination in U.S. Housing Markets." *Housing Policy Debate* 3(2):685-745.

Feins, Judith. 1976. "Urban Housing Disinvestment and Neighborhood Decline: A Study of Public Outcomes." Ph.D. diss., University of Chicago.

Fernández Kelly, M. Patricia. 1983. *For We Are Sold, I and My People: Women and Industry in Mexico's Frontier.* SUNY Press.

Galster, George C. 1992. "Research on Discrimination in Housing and Mortgage Markets: Assessment and Future Directions." *Housing Policy Debate* 3(2):639-83.

García, F. Chris. 1988. *Latinos and the Political System.* University of Notre Dame Press.

Harrison, Bennett, and Barry Bluestone. 1988. *The Great U-Turn: Corporate Restructuring and the Polarizing of America.* Basic Books.

Hero, Rodney E. 1992. *Latinos and the U.S. Political System. Two-Tiered Pluralism.* Temple University Press.

Hoyt, Homer. 1933. *One Hundred Years of Land Values in Chicago.* University of Chicago Press.

Hughes, Elizabeth Ann. 1925. *Living Conditions for Small Wage Earners in Chicago.* Chicago: Department of Public Welfare.

____. 1937. *Illinois Persons on Relief in 1935.* WPA Project No. 165-54-6018. Chicago: Public Works Administration.

Jackson, Peter. 1981. "Paradoxes of Puerto Rican Segregation in New York." In *Ethnic Segregation in Cities,* edited by Ceri Peach, Vaughn Robinson, and Susan Smith. Croom Helm.

Jennings, James, and Monte Rivera. 1984. *Puerto Rican Politics in Urban America.* Greenwood Press.

Jiménez, José Cha. 1974. "Statement to the Spanish Speaking Peoples Study Commission of the State of Illinois." Chicago: Spanish-Speaking Peoples Study Commission Hearings, 15 November.

Jones, Anita. [1928] 1971. *Conditions Surrounding Mexicans in Chicago.* R&E Research Associates.

Kirstein, Peter N. 1977. *Anglo over Bracero: A History of the Mexican Worker in the U.S. from Roosevelt to Nixon.* R&E Research Associates.

Kiser, George C., and Martha W. Kiser (eds.). 1979. *Mexican Workers in the U.S.: Historical and Political Perspectives.* University of New Mexico Press.

Kushner, James E. 1992. "Federal Enforcement and Judicial Review of the Fair Housing Amendments Act of 1988." *Housing Policy Debate* 3(2):537-99.

Lane, James B., and Edward J. Escobar (eds.). 1987. *Forging a Community: The Latino Experience in Northwest Indiana, 1919-75.* Cattails Press.

Latino Institute. 1983. *Latinos in Metropolitan Chicago: A Study of Housing and Employment.* Chicago: Latino Institute.

____. 1993. "Suburban Latinos: The Working Poor?" Press release, 3 May.

Luttrell, J. 1970. "Public Housing and Integration: A Neglected Opportunity." *Columbia Journal of Law and Social Problems* 6:253-79.

Maldonado, Edwin. 1987. "Contract Labor and the Origins of Puerto Rican Communities in the U.S." Pp. 201-12 in *Forging a Community: The Latino Experience in Northwest Indiana, 1919-75,* edited by Edward J. Escobar and James B. Lane. Cattails Press.

Martínez, Manuel. 1989. *Chicago: Historia de Nuestra Comunidad Puertorriqueña, Photographic Documentary.* Reyes.

Massey, Douglas S. 1979. "Effects of Socioeconomic Factors on the Residential Segregation of Blacks and Spanish-Americans in U.S. Urbanized Areas." *American Sociological Review* 44:1015-22.

____. 1987. "Ethnic Residential Segregation: A Theoretical Synthesis and Empirical Review." *Sociology and Social Research* 69:315-50.

Massey, Douglas S., and Nancy Denton. 1987. "Trends in the Residential Segregation of Blacks, Hispanics, and Asians: 1970-80." *American Sociological Review* 52:802-25.

____. 1988. "Suburbanization and Segregation in U.S. Metropolitan Areas. *American Journal of Sociology* 94:592-626.

____. 1993. *American Apartheid: Segregation and the Making of the Underclass.* Harvard University Press.

Meléndez, Edwin. 1988. *Labor Market Structure and Wage Inequality in New York: A Comparative Analysis of Hispanics, Non-Hispanic Blacks and Whites.* MIT, Department of Urban Studies and Planning.

Meléndez, Edwin, C. Rodríguez, and J.B. Figueroa (eds.). 1991. *Hispanics in the Labor Force: Issues and Policies.* Plenum.

Morales, Rebecca, and Frank Bonilla. 1993. *Latinos in a Changing U.S. Economy.* Sage.

Nathan, Debbie. 1984. "On Panic and Puerto Ricans: Is Blockbusting Back on the South Side?" *Chicago Reader,* 8 June.

Omi, Michael, and Howard Winant. 1986. *Racial Formation in the U.S.: From the 1960s to the 1980s.* Routledge & Kegan Paul.

Orfield, Gary, and Ken Gaebler. 1991. *Residential Segregation and the 1990 Census.* Chicago Urban League.

Pachón, Harry. 1985. "Political Mobilization in the Mexican American Community." In *Mexican Americans in Comparative Perspective,* edited by Walker Connor. Urban Institute.

Padilla, Elena. 1947. "Puerto Rican Immigrants in New York and Chicago: A Study in Comparative Assimilation." Ph.D. diss., University of Chicago.

Padilla, Felix M. 1987. *Puerto Rican Chicago.* University of Notre Dame Press.

Park, Robert E., and Ernest W. Burgess. 1921. *Introduction to the Science of Sociology.* University of Chicago Press.

Park, Robert E., Ernest W. Burgess, and Roderick D. McKenzie. 1925. *The City.* University of Chicago Press.

Pax, Frank X. 1948. "Mexican Americans in Chicago: A General Survey." Metropolitan Welfare Council Papers, Chicago Historical Society, Box 147.

Peterman, William, and Qi Sanshi. 1991. "Lending Discrimination in Metropolitan Chicago." Pp. 21-47 in *Credit by Color*. Chicago Area Fair Housing Alliance.

Portes, Alejandro, and John Walton. 1981. *Labor, Class, and the International System*. Academic Press.

Puente, Sylvia, and Alejandrina Vásquez. 1989. "The Condition of Latino Housing in Chicago." Master's thesis, University of Chicago.

Redfield, Robert. 1924-25. "Mexicans in Chicago 1924-25" (Diary). Robert Redfield Papers, Special Collections. University of Chicago.

Reisler, Mark. 1976. *By the Sweat of Their Brow: Mexican Immigrant Labor in the U.S., 1900-40.* Greenwood Press.

Rosales, Francisco, and Daniel T. Simon. 1978. "Los Trabajadores Chicanos en la Industria Siderúrgica y el Sindicalismo en el Medio Oeste de 1919 a 1945." Pp. 146-57 in *Origenes del Movimiento Obrero Chicano*, edited by Juan Gómez Quiñones and Luis E. Arroyo. Mexico, D.F.: Ediciones Era.

Sánchez-Korrol, Virginia. 1983. *From Colonia to America: The History of Puerto Ricans in New York.* Greenwood Press.

Sassen, Saskia. 1988. *The Mobility of Labor and Capital*. Cambridge University Press.

Shevky, Eshref, and Wendell Bell. 1955. *Social Area Analysis: Theory, Illustrative Application, and Computational Procedures*. Stanford University Press.

Shlay, Anne B. 1987. "Credit on Color: Segmentation, Racial Transition, and Housing Credit Flows." In *Fair Housing in Metropolitan Chicago: Perspectives after Two Decades*, edited by Gary Orfield. Chicago Area Fair Housing Alliance Research Project.

____. 1988. "Not in That Neighborhood: The Effects of Population and Housing on the Distribution of Mortgage Finance within the Chicago SMSA." *Social Science Research* 17:137-63.

Smith, Neil. 1979. "Gentrification and Capital: Practice and Ideology in Society Hill." *Antipode* 11:24-35.

Squires, Gregory D., Larry Bennett, Kathleen McCourt, and Philip Nyden. 1987. *Race, Class, and the Response to Urban Decline*. Temple University Press.

Squires, Gregory D., Ruthanne DeWolfe, and Alan S. DeWolfe. 1979. "Urban Decline or Disinvestment: Uneven Development, Redlining, and the Role of the Insurance Industry." *Social Problems* 27:79-95.

Suttles, Gerald D. 1972. *The Social Construction of Communities*. University of Chicago Press.

Taylor, Paul. 1932. *Mexican Labor in the U.S.: Chicago and the Calumet Region*. University of California Press.

Teitelbaum, Michael S. 1985. *Latin Migration North: The Problem for U.S. Foreign Policy*. Council on Foreign Relations.

Tobias, Ruth Anne, and Chandrima Roy. 1993. *Diversity in the Northwest Suburbs*. Center for Governmental Studies, Northern Illinois University.

Turner, Margery A., Raymond J. Struyk, and John Yinger. 1991. *Housing Discrimination Study. Synthesis*. Urban Institute and Syracuse University.

Urban Institute. 1991. *U.S. Housing Discrimination Survey*. Urban Institute.

USFHA. 1938. *Underwriting Manual*. Government Printing Office.

____. 1947. *Underwriting Manual*. Government Printing Office.

Valdez, Avelardo, and Richard C. Jones. 1984. "Geographical Patterns of Undocumented Mexicans and Chicanos in San Antonio, Texas: 1970 and 1980." Pp. 218-35 in *Patterns of Undocumented Migration: Mexico and the United States*, edited by Richard C. Jones. Rowman & Allenheld.

Wachter, Susan M., and Isaac F. Megboulegbe. 1992. "Racial and Ethnic Disparities in Homeownership." *Housing Policy Debate* 3(2):333-70.

Weink, Ronald E. 1992. "Discrimination in Urban Credit Markets: What We Don't Know and Why We Don't Know It." *Housing Policy Debate* 3(2):17-40.

White, Michael J. 1987. *American Neighborhoods and Residential Differentiation.* Russell Sage Foundation.

Yinger, J. Milton. 1985. "Assimilation in the U.S.: The Mexican Americans." Pp. 29-55 in *Mexican Americans in Comparative Perspective,* edited by Walker Connor. Urban Institute Press.

Segregation and Crime: The Effect of Black Social Isolation on the Rates of Black Urban Violence*

EDWARD S. SHIHADEH, *Louisiana State University*
NICOLE FLYNN, *Louisiana State University*

Abstract

Prior segregation-crime research has failed to recognize that segregation has many geographic forms and each may have a distinct macrosocial path to crime. We sharpen the conceptual link between segregation and crime by considering how the social isolation of urban blacks increases black violence. Using race-disaggregated Uniform Crime Report (UCR) and census data for 1990, we examine the link between black social isolation and the rates of black homicide and robbery in U.S. cities. In contrast to previous research, which employs the index of dissimilarity (D) as a default indicator of segregation (which it is not), we measure the spatial isolation (P) of blacks from whites. Black isolation emerges as a strong predictor of the rates of black violence in major U.S. cities, a finding that may account for prior evidence of a link between segregation and violence at the macro level.

High levels of violence in urban black communities are a widely acknowledged though often neglected research issue. While homicide rates in the U.S. are among the highest in world, of particular concern is this country's disproportionately high rates of minority crime. For instance, while blacks comprise only about 12% of the U.S. population (Current Population Reports 1988), they account for over one half of homicides and about two-thirds of robbery arrestees (U.S. Dept. of Justice 1988). About 50% of all black urban males are arrested for a serious crime at least once, compared to only 14% of white urban males, and blacks are six times more likely to spend time in prison or jail (Petersilia 1985).

There is a growing concern that high rates of violence in black communities may be related to black-white residential segregation in urban areas. Indeed, while minority segregation in the U.S. cities is pervasive, black spatial segregation is particularly entrenched and appears to be highly resistant to change

* *We are grateful to Katharine Donato, Scott Feld, Daniel Lichter, Roger Wojtkiewicz, and the anonymous reviewers for helpful comments.* *Direct correspondence to Edward S. Shihadeh, 126 Stubbs Hall, Department of Sociology, Louisiana State University, Baton Rouge, LA 70803*

 Social Forces, June 1996, 74(4):1325-1352

(Farley & Frey 1994; Massey & Mullan 1984; Massey 1990; Massey & Denton 1988, 1993). One especially pernicious outcome of black-white segregation is in how it socially isolates many urban blacks from mainstream society. Recent research on the rise of the urban underclass and the ghettoization of black inner-city areas suggests that the geographic isolation of low-income urban blacks from white-dominated neighborhoods has anchored many black residents to areas of multiple disadvantage (e.g., Massey & Denton 1993; Wilson 1987, 1993). This creates a significant barrier to the social mobility of many urban blacks in the U.S. and may contribute to the rates of violence in black communities.

While the handful of studies that have examined the relationship between segregation and crime have been instructive, they contain several important limitations. First, they have relied heavily on the notion that segregation is an ascriptive form of inequality that generates high crime rates by creating hostile sentiments and/or a weakened attachment to social norms restraining the use of violence. Unfortunately, this approach ties the study of segregation and crime to a body of research that has provided mixed and inconsistent evidence of a link between inequality and crime (see review below). It also assumes that the crime-producing mechanism of segregation is the same as it is for other forms of inequality (e.g., income inequality, educational inequality). However, recent research on segregation and the social isolation of the underclass suggests that this may not be the case.

Second, previous studies view segregation in only one way — as the degree of spatial *unevenness* in the distribution of blacks versus whites. However, segregation has a number of empirically distinct distributional forms (Massey & Denton 1988). The exclusive focus on spatial unevenness as a default indicator of segregation fails to capture other theoretically relevant dimensions of residential segregation — namely the social isolation of blacks from whites. And third, while several studies disaggregate crime rates by race, rarely are they concerned with measuring the direct impact of segregation on crimes committed by blacks. This raises a concern since both segregation and its possible outcomes, such as serious crime, are problems faced mainly by black communities.

We extend the research on segregation and crime by suggesting an alternative causal mechanism by which segregation generates violence. Our objective is to examine the impact of one specific form of segregation, black isolation from whites, on the rates of black violence in U.S. cities for 1990. We begin by discussing the theoretical link between black isolation and black urban violence. A brief simulation is then presented in order to clarify the distinction between black-white spatial unevenness — which is the more commonly used dimension of segregation — and black isolation from whites. We then estimate several models predicting racially disaggregated rates of homicide and robbery, using an indicator of black isolation and a variety of racially disaggregated predictors. Our analysis concludes with a discussion of the theoretical and methodological implications of the findings.

Black Residential Segregation and Crime

Though racial discrimination in the sale and rental of housing was banned by the Civil Rights Act of 1968, an enduring feature of urban areas in the U.S. is the extreme residential segregation of blacks. Despite the liberalization of white attitudes, endorsing the principle of integration has been easier than accepting black neighbors (Farley & Frey 1994). During the 1980s, the average black-white index of dissimilarity in major metropolitan areas was exceedingly high and declined only modestly, from an average of 69 to 65 from 1980 to 1990. Massey and Denton (1993) argue that the source of black segregation is in a legacy of white discrimination against blacks, producing a physical separation of the races and an isolation of many blacks from mainstream residential communities and institutions. Blacks are discouraged from settling in white neighborhoods in order to preserve the racial character of the area. Alternatively, whites abandon communities with a growing black population, producing a line of retreating white neighborhoods in the face of black residential expansion. In any case, blacks lack equal opportunity in housing and are more likely to be tied to communities with multiple disadvantages, where the rates of unemployment, poverty, school dropout, family disruption, and teen pregnancy are disproportionately high. By blocking channels to social and economic mobility and undermining the legitimacy of established social norms, high rates of residential segregation are presumed to generate various forms of deviance and violent behavior.

While evidence suggests that residential segregation has a number of negative outcomes, comparatively little is known about its relationship with respect to crime (Logan & Messner 1987; Messner & South 1986; Peterson & Krivo 1993; Potter 1991; Rosenfeld 1986; Sampson 1985; South & Felson 1990). Messner and South (1986) and South and Felson (1990) find that segregation tends to increase rates of *intra*-racial victimization, but decrease rates of *inter*-racial victimization. Logan and Messner (1987) examine the impact of segregation on violence rates in the suburbs of large urban areas and report a positive relationship between segregation and murder for 1980, but not for 1970. Rosenfeld (1986) finds an empirical association between homicide rates and segregation in large U.S. metropolitan areas in 1970. Among those that directly test the impact of segregation on black crime, Sampson (1985) reports the unexpected finding that segregation predicts homicide rates among whites but not blacks. In contrast, Potter (1991) finds that segregation is related to the black-white difference in homicide rates, and Peterson and Krivo (1993) find that segregation predicts stranger and acquaintance homicide among urban blacks in 1980.

While these studies are instructive, they are problematic and thereby limit a thorough understanding of how the residential segregation of blacks generates high crime rates. First, there has been a heavy reliance on the notion that segregation as an ascriptive form of inequality generates violence by causing frustration and aggression and by weakening the attachment to conventional normative structures. In the most prominent statement of that thesis, Blau and Blau (1982) use a frustration-aggression framework to argue that inequality based on ascriptive characteristics such as race can lead to hostile sentiments

which, in extreme cases, find expression in the form of violence. Extending Merton's work on social structure and anomie, Logan and Messner (1987) have argued that inequality is a system disjuncture that weakens the legitimacy of social norms and reduces the capacity of communities to exert social control.

Unfortunately these arguments tie the analysis of segregation to a body of research that has produced mixed results and conflicting evidence regarding a link between racial inequality and serious violence. Our audit of the literature reveals that while some find support that inequality directly increases crime rates (Blau & Golden 1986; Blau & Schwartz 1984), most find either partial support (Balkwell 1990; Messner & Golden 1992; Sampson 1985; Williams 1984) or no support at all (Balkwell 1983; Carroll & Jackson 1983; Shihadeh & Steffensmeier 1994; Harer & Steffensmeier 1992; Patterson 1991; Rosenfeld 1986). We agree that segregation is one dimension of a racially stratified system. What troubles us is the underlying assumption that the crime-producing mechanism of black-white residential segregation is the same as that identified for other types of racial inequality (e.g., inequality by income, education, occupational prestige, etc.). Perhaps the research on segregation and crime can be enhanced by identifying specific causal paths by which the geographic separation of the races leads to high rates of minority crime.

Second, previous segregation-crime research has examined only one form of segregation, the spatial *unevenness* between blacks and whites. Typically calculated using the index of dissimilarity (*D*), urban blacks are considered segregated if their residences are unevenly distributed across areal units with respect to those of whites. But Massey and Denton (1988) identify five forms of black-white segregation and each is conceptually and empirically independent of the other (unevenness, isolation, centralization, concentration, and clustering).[1] These are not five measures of the same phenomenon, but five distinct ways in which two or more groups live separately from one another . Thus, the practice of using unevenness (*D*) as the generic form of segregation, rather than as merely *one* form, is unwarranted since unevenness in spatial distribution does not represents all forms of segregation. Moreover, previous research has not provided a rationale as to why unevenness, versus other forms of segregation, is the most appropriate match for the theories under examination. Perhaps a more informed understanding of how segregation engenders violence requires that we focus on one particular distributional characteristic of black-white segregation and then identify a specific theoretical mechanism linking it to violence.[2]

Finally, with the exception of Sampson (1985) and Peterson and Krivo (1993), previous studies have not examined the direct impact of segregation on *black* violence. Yet it is reasonable to expect that the negative consequences resulting from segregation are directed mainly towards the minority in question, not the population as a whole. Since blacks bear the brunt of the adverse effects of black-white residential segregation, research should focus more closely on the specific outcomes to their communities. Indeed, recent evidence suggests that the macrosocial paths leading to high rates of black urban violence differ from those contributing to high rates of white urban violence (e.g., Harer & Steffensmeier 1992; Huff-Corzine, Corzine & Moore 1986; LaFree, Drass & O'Day 1992; Sampson 1987; Shihadeh & Steffensmeier 1994). Evidence also suggests that

crime models that use nonracially disaggregated data mask potentially important race differences in the sources of urban violence.

In brief, the evidence indicates that segregation is an embedded feature of urban areas in the U.S. and may have a variety of negative outcomes on the lives of many black residents. The results from several studies suggest that black-white residential segregation contributes to the rates of violent crime, particularly among blacks. However, the knowledge of the theoretical and causal nature of this relationship remains very limited. Our objective is to contribute to our understanding about the consequences of segregation and the macrosocial paths leading to black violence by considering recent research on the rise of the black urban underclass.

BLACK SOCIAL ISOLATION

Among the various ways that blacks can be segregated from whites, one particular form, *black isolation*, exists when the degree of potential contact between blacks and whites is low. Because of perpetuating institutional arrangements and individual actions, blacks face considerable difficulty in accessing nonminority neighborhoods. Blacks are excluded from white areas either through subtle means, such as the denial of mortgages, or by direct threat and intimidation. Conversely, whites are inclined to move when their preference for limited interracial contact is threatened by an influx of blacks (Farley et al. 1978; Farley et al. 1994). The resulting limited contact between blacks and whites may isolate many blacks from the rest of society and severely limit their chances for social mobility (Massey & Denton 1993; Massey & Mullan 1984; Wilson 1987).

Although other groups have experienced isolation, it is a particularly serious problem in urban black communities. First, blacks in the U.S. experience an isolation that is unprecedented and qualitatively different from that faced by other ethnic groups (Massey & Denton 1993). Although certain parts of a city may have been publicly identified with a particular immigrant-ethnic group, such as Italian or Polish, these areas, even during the height of immigration, have never shown the degree and persistence of isolation found in low-income black communities (Massey & Denton 1993; Philpott 1978). Indeed, among the ten largest U.S. cities in 1980, the average black lived in a neighborhood that was in excess of 80% black, and the vast majority resided in areas that are were 100% black (Massey & Denton 1988). Second, black isolation not only isolates blacks themselves, but also the negative characteristics that are collinear with black urban life. Black isolation combines the weight of poverty, joblessness, welfare dependency, teenage childbearing, and other indicators of social malaise and concentrates them geographically in black neighborhoods (Wilson 1993). Social isolation is a problem for many black-dominated areas not because these communities have some intrinsic weakness in their adaptation to negative structural conditions, but because these unprecedented levels of isolation converge a set of multiple disadvantages into a single ecological space.

In broad terms, spatial isolation leads to social isolation, which in turn has a variety of negative outcomes including serious crime. Below, we discuss a number of ways how black isolation can lead to the economic, cultural, and

political decline of urban black communities. The logic of our argument is not hinged on any single reason, but on a combination of following possible consequences of black isolation. Our analytical objective is to establish whether a cross-sectional relationship exists between black social isolation and serious black crime across major cities in the U.S. However, for background and theoretical clarity we recognize in the discussion below that the relationship between black isolation, the structural characteristics of urban areas, and serious crime is processual in nature — as is the relationship between other social factors.

Economic Factors

One of the major concerns of black isolation is that it may undermine the attachment of low-income blacks to the labor force. Unless residents of low-income black areas work outside the neighborhood, they are unlikely to have sustained contact with anyone else who is not black and who has steady employment (Pedder 1991; Yancey 1985). The effects of this structural constraint loom large when one considers that in central cities 54% of black men aged 18-29 are either out of the labor force, unemployed, or underemployed (Lichter 1988). Massey and Denton (1993) lament that the residents of these neighborhoods "necessarily live within a very circumscribed and limited social world. They rarely travel outside of the black enclave, and most have few friends outside of the ghetto . . . Because blacks have weak links to white society, they are not connected to the jobs that white society provides" (161-62). This does not imply that ghetto residents have a reduced willingness or motivation to work. Rather, with limited access to job information systems and a lack of successful role models, many blacks are disadvantaged in the competition for employment, particularly given the shrinking base of low-skill jobs in central cities (Kasarda 1990). The inevitable result is a weakened labor force attachment among large segments of center-city blacks that "raises the likelihood that [*they*] will turn to illegal or deviant activities for income, thereby further weakening their attachment to the legitimate labor market" (Wilson 1993:21).

An underlying source of black isolation may be the loss of the black middle class to the more affluent areas in the suburbs, which removed the vertical integration of many center-city black communities and transformed them into areas of concentrated poverty (Wilson 1987, but see Massey, Gross & Shibuya 1994). While black suburbanization may have contributed to the economic decline of black inner-city neighborhoods, a contributing factor may be a downward mobility among middle-class stayers in these areas. Residents who remain behind find that one of the most important sources of middle-class capital accumulation, housing value, is steadily eroded. As the neighborhood declines, housing stock deteriorates, and institutional decisions by banks and insurance companies make it difficult to acquire reasonable mortgages and insurance (Bradford & Rubinowitz 1975; Urban-Suburban Investment Study Group 1975), a practice known as "redlining." Signs that the neighborhood is declining create a self-fulfilling prophecy that leads to a downward spiral in property values and a corresponding decline in the wealth of anchored residents (Skogan 1992). Thus those left behind in declining areas will experience downward social mobility as home equity disappears. Moreover, if home equity

is a conduit to other sources of capital, such as auto and consumer financing, or college and small-business loans, structural declines in home value could further distance local residents from mainstream economic and social opportunities.

Cultural

If members of ghetto communities are influenced by the beliefs, modes of behavior and the experiences of relevant others around them, black isolation may produce an inversion of mainstream values that further increases the social disability of blacks. This is not a culture of poverty perspective, which focuses on negative subcultural and personality traits that acquire a life of their own and continue to influence behavior even when structural conditions improve (Banfield 1974; Curtis 1975; Gastil 1971; Lewis 1961, 1968). Rather, we emphasize how oppositional ghetto practices are proximate outcomes to a structural condition of social isolation. For instance, Elijah Anderson (1993, 1994) uses ethnographic research to link certain ghetto-specific practices to black social isolation. He maintains that because many young blacks are isolated from the white economy and white job networks, they have little hope of acquiring meaningful employment sufficient to support a family. They are deprived, therefore, of the primary conduit by which adult men in mainstream society prove their worth. Instead they turn to male peers for support and assert their manhood by fostering a self-image based on violence. Anderson argues that ghetto-specific practices, such as overt sexuality, gang membership, and open displays of aggression reflect the emergence of an alternative status system that legitimates behaviors that would be held in contempt by the wider society. He further notes that clothing, jewelry, hair styles, facial expressions, and gait are all used to shore up an identity of physical toughness. Even young men with otherwise nonviolent predispositions feel compelled to assume a street-oriented demeanor, emphasizing a distinctive linguistic pattern (Labov 1980) and the willingness to use violence (Anderson 1994).

Anderson (1990) also emphasizes how social isolation in black ghettos can invert conventional values about childbearing. Traditional standards stress that pregnancy should follow marriage, or vice versa, and should be accompanied by financial stability. But the female subjects he interviewed in ghetto areas appeared to reject that notion. On the contrary, a conduit to recognition and prestige among girls and young women in the ghetto appeared to be childbearing itself, regardless of the trappings of marriage or cohabitation and financial stability. This ideology is countercultural in that those women who have children have acquired adult status in the eyes of peers, while those without children have not. But since mother-only households make up a disproportionate share of the poor, this leads to legitimate concerns about the intergenerational transfer of poverty (McLanahan & Booth 1989). Much of the extant crime research has generally failed to yield evidence of a significant individual-level relationship between family structure and crime — that single-parent households are more likely to produce juvenile delinquents (Ensminger, Kellam & Rubin 1983; LaFree, Drass & O'Day 1992; Ross & Sawhill 1975; but see Matsueda & Heimer 1987). Yet at the aggregate level, Sampson (1987) and Shihadeh and Steffensmeier (1994) find extremely strong associations between the rates of black female-headed household and the rates of black violence. They

argue that when single-parent households are the norm, this weakens the formal and informal capacity of communities to exert social control over their members which in turn leads to higher rates of offending (see Shihadeh & Steffensmeier 1994). In any case, adaptations of household structure within isolated black communities represents a potentially important component of the segregation-crime link.

Black isolation further distances black residents from mainstream society by removing the stigma of economic and social failure, and weakens the capacity of major institutions to socialize and control their members. For instance, it can devalue education and even impose a direct social pressure for educational failure (Crain & Mahard 1978; Fordham & Ogbu 1986). Many bright and motivated blacks in segregated low-income areas face enormous pressure not to succeed in school for fear of being labeled as too "white" by other kids (Fordham & Ogbu 1986). Worst of all, the social stigma of incarceration — perhaps society's harshest form of social control — is weakened. Brown (1984) suggests that a significant portion of young male ghetto residents view doing time in prison as a form of graduate school and a rite of passage to peer acceptance.

Political

Black isolation from whites may also lead to political and economic disenfranchisement of black neighborhoods (see discussion in Massey & Denton 1993). Historically, ethnically integrated neighborhoods formed the basis for powerful political coalitions (Dahl 1961, 1967; Pinderhughes 1987). Because no single ethnic group typically dominated white areas (Massey & Denton 1993), resources allocated to any specific ethnic group is of benefit to all members of the local community. Given this interweaving of interests, political representatives have to pay attention to the demands of a wide spectrum of voters, irrespective of ethnic affiliation (Pinderhughes 1987). The geographic isolation of blacks, however, undermines the power of pluralist politics and means that whites have less to gain when public funds and resources are channeled to all black areas and away from multiethnic communities (Edsall & Edsall 1991; Grimshaw 1980; Katznelson 1976; Kemp 1986; Spear 1967). Black isolation further contributes to the disinvestment and decline in many urban black areas by making them vulnerable to cutbacks in public funds and services. Rather than risk antagonizing white voters, white politicians are more apt to target black neighborhoods for spending cuts because of the minimal political fallout (D. Wallace 1990; R. Wallace 1991).

Limitations in the political influence of socially isolated black areas may reduce the effectiveness of traditional forms of neighborhood-level political expression. Social disorganization theorists emphasize how a network of informal and formal social controls maintain desirable conduct and conventional lines of action in a community (Bursik & Grasmick 1993; Kasarda & Janowitz 1974; Shaw & McKay 1931). But this type of control is *internal* to the community. In contrast, a public level of control concerns the ability of a community to control the *external* environment in order to procure government aid and services, as well as the passage or defeat of laws and ordinances favorable to the community (Hunter 1985). In general terms, public control refers to the use

of external political, social, or economic structures — such as local governing bodies — for the benefit of the local community (Bursik & Grasmick 1993). This public control embodies the capacity of the community to manipulate external resources in order to defend its interests (Hunter 1974). While membership in civic associations, churches, and participation in local politics is a conduit of this expression, its effectiveness in isolated ghetto areas may be very limited. Evidence suggests that residents of low-income black areas are less likely than otherwise-comparable blacks to be members of churches and local organizations and are less active in electoral politics (Brown 1991). Furthermore, lack of participation in community-based initiatives is a predictor of both adult and juvenile crime rates (Sampson & Groves 1989). Thus, in depriving blacks of political empowerment, social isolation may contribute to their economic and social decline.

While black isolation from whites may have a number of potentially harmful effects on the economic, cultural, and political structures in black communities, in a broader sense it can also be tied to the emerging ideas on social capital. Social capital refers to the structure of interpersonal relations and institutional linkages that facilitate effective ties that bind a person to the community or society and that enhance an individual's life chances. "Social capital is productive, making possible the achievements of certain ends that in its absence would not be possible" (Coleman 1988:98). Thus, social capital draws a strong connection between community-level processes and the lives of individuals. "Entire communities and countries, as well as the individuals within them, are locations of the accumulation of social capital" (Hagan 1994:69).

If social capital refers to the features of structured groupings that contribute to productive lives and conventional behavior, then those who lack access to it are at a disadvantage in seeking upward mobility (Bourdieu 1986). As a structured form of inequality, black isolation may reduce the accumulation of social capital in black communities for a number of reasons. If Bourdieu is correct — if economic capital is at the root of all forms of capital, including social capital — then the negative economic effects of black isolation will reduce the accumulation of social capital in black communities. But even from a cultural standpoint, communities extend social capital to members by cultivating values such as a work and family ethic, by instilling in males a sense of parental responsibility for children, and by creating confidence in the returns from education. These elements are critical for the transition from youth to successful adulthood in mainstream life, but social isolation may create conditions that hinder the capacity of black communities to extend these essential forms of social capital to their members.

In the following analysis, we go beyond previous segregation-crime research and examine the impact of black isolation on the rates of black urban violence. Initially, we run a brief simulation using hypothetical data in order to clarify the distinction between our measure, black isolation from whites, and the measure used in previous research, unevenness in spatial distribution. We then estimate models that incorporate racially disaggregated data for U.S. cities in 1990.

Data and Methods

The units of analysis in this study are the 151 U.S. cities with populations of 100,000 or more, and with at least 5,000 blacks in 1990. Socioeconomic data for these cities are obtained from Summary Tape File 3C of the 1990 census. Segregation indices for cities are derived using tract-level data from Summary Tape File 1A. Data on black political representation within cities are manually entered from publications of the Joint Center for Political Studies (1989) and Carpenter 1992. Since race-disaggregated arrest data for central cities are not available on publicly released Uniform Crime Report (UCR) periodicals, we obtained them through a special tabulation provided to us by the Federal Bureau of Investigation. We derive black homicide and robbery arrest rates for 1990 by first averaging the city-level arrest counts over a three year period — 1989, 1990, 1991 — to minimize any year-to-year fluctuations in serious crime. We then divide that average by the total city population aged 10 and older. We exclude those aged 0-9 for two reasons. First, children in this age group are unlikely to be a part of the arrest statistics for homicide and robbery. Second, since the 0-9 age group comprises a substantial part of the U.S. population, any variation across cities in the relative size of that group could potentially bias the results. Finally, because these rates are heavily skewed, we transform them to their natural log in order to induce normality and to avoid heteroskedasticity.[3]

Official data have been criticized on the grounds that they measure the reaction of law enforcement agencies rather than criminal behavior itself. However, for very serious crimes such as homicide and robbery, UCR arrest data appear to reflect accurately the rates of black and white offending (Hindelang 1978; Laub et al. 1987). Seriousness of the crime is the strongest predictor of arrest (Black & Reiss 1970; Gottfredson & Gottfredson 1980; Gove, Hughes & Geerken 1985). Crime severity accounts for the greatest variance in whether a crime becomes officially recorded (Gove, Hughes & Geerken 1985). Based on a comparison of victimization data from the National Crime Survey and arrest data from the UCR, Hindelang (1978) concludes that there is little or no race-bias in robbery arrests. LaFree, Drass and O'Day (1992) compare arrest rates with total crimes known to the police for the years 1957-88 and find correlations of .98 and .97 for robbery and murder, respectively. Hence the available evidence indicates that homicide and robbery arrest rates accurately reflect offending in the total population and among blacks and whites. In the multivariate analysis, we predict these violence rates using ordinary least squares regression models.

SEGREGATION: BLACK ISOLATION FROM WHITES

Isolation refers to the degree of potential contact between minority and majority members across geographic subareas of a spatial unit. Specifically, we focus on *black isolation* (P) from whites within a city. We measure black isolation using Lieberson's (1981) interaction index, which reflects the probability, ${}_bP^*_w$, that a randomly drawn black in the city interacts with a white: This is defined as

$$ {}_bP^*_w = \sum_i \beta_i(w_i/t_i) $$

where B_i is the proportion of all blacks in the city that are located in tract *i*, and w_i and t_i are the number of whites and total population, respectively, in tract *i*. Values range from 0 to 1.0; the former indicating complete spatial isolation of blacks from whites, and the latter reflecting certainty that a given black shares a tract with whites. Massey and Denton (1988b) demonstrate that ${}_bP^*_w$ is the preferred measure of interracial exposure because it loads highly on a single dimension in a factor analysis and has a straightforward interpretation. So as to avoid switching back and forth between the term *isolation* and *interaction*, we reverse the sign of ${}_bP^*_w$ so that we may consistently refer to this as a measure of isolation.

Given the reliance of previous segregation-crime research on unevenness as a measure of segregation, we include an indicator of it in our models. *Unevenness* (*D*) refers to the difference between two groups in their geographic distribution across areal units. For instance, segregation exists if blacks are located in relatively few areas of a city and whites exhibit a comparatively uniform spatial distribution. Typically calculated using the index of dissimilarity, unevenness is defined as:

$$D = [(1/2)\,|B_i - W_i|] * 100$$

where B_i refers to the proportion of all blacks in the city who live in tract *i*, while W_i has a similar meaning for whites. *D* varies from 0 to 100 and is interpreted as the proportion of minority residents that would have to change tracts to produce a uniform race distribution across the city (Shryock & Siegal 1976).

MEASUREMENT OF INDEPENDENT VARIABLES

In our theoretical discussion, we identify three process — economic, cultural, and political — through which black isolation could potentially raise the rates of black violence. Accordingly, we group the independent variables into the same three categories, plus a fourth for standard controls. We recognize that individual predictors may reflect overlapping processes because economic variables, for instance, might be related to cultural responses and/or political conditions and vice versa. However, if the relationship between black isolation and black violence operates via economic, cultural, and political factors, then including indicators of these in our models should weaken the link, if any, between black isolation and black violence.

We include three predictors that reflect black socioeconomic status in central cities. The first is the rate of *black employment*, measured as the proportion of all working-age blacks (16-64) in the city that are employed. Calculating our measure of black employment from only those in the labor force can overestimate black employment since many blacks exit the labor force as discouraged workers (Wetzel 1995). Evidence shows that joblessness increases black violence both directly (Allan & Steffensmeier 1987; Devine, Sheley & Smith 1988; Shihadeh & Ousey 1996) and indirectly as well (Sampson 1987; Shihadeh & Steffensmeier 1994). Second, we include *black poverty*, the proportion of black residents in the central city whose 1989 income fell below the poverty line. As

an indicator of home equity, we measure *black renters* as the proportion of black residents who rent rather than own their residence.

Two variables that measure potential differences in culture are *black youth attachment* and *female-headed households*. The first is the proportion of black teens aged 16 to 19 that are unemployed or not in the labor force, not in school, not in the armed services, and have no high school degree. These are individuals that have either no attachment or weak attachment to major social institutions such as the labor market, armed services, high school, as well as vocational or other secondary educational institutions. In the objectively harsh environments within many black urban areas — where the rates of joblessness, and persistent and intergenerational poverty are high and where legitimate opportunities are lacking but welfare use is common — a widespread sense of cynicism and frustration may make it difficult for the community to cultivate allegiance to mainstream goals. Should the residents of these communities view the returns to education as poor and the prospects of adequate and meaningful employment as slim, mainstream goals such as securing a high school degree or postsecondary degree or certificate as well as looking for or getting a job may not hold the same priority. In other words, the simultaneous combination of a youth aged 16 to 19 being jobless or not in the labor force, without a diploma, not currently enrolled in school, and not in the armed forces strongly suggests a marginalization or detachment from a normative social order. This detachment is of particular concern among those aged 16 to 19 since they are within the crime-prone age group for serious violence.

The extensive discussion about the racial differences in the rates of female-headed households suggests that high rates of female-headed households in black communities are due to economic factors: High rates of black male joblessness have reduced the economic incentives to marriage for black women, elevating the rates of black female-headed households (Wilson 1987). But equally compelling is the suggestion that the source of racial differences in household structure are cultural in nature: Because of historical circumstances and contemporary structural arrangements, black women are more likely to choose female-headed households as the preferred living arrangement. This may be a cultural difference visible more than a century ago (Cherlin 1992; Gutman 1976; Morgan et al 1990; Sudarkasa 1981). But as our earlier discussion of Anderson's (1990) ethnographic research suggests, normative adaptations to the social isolation of the ghetto may exacerbate the racial differences in family structure. To recall, he finds that a contributing factor to the high rates of mother-only families in ghetto areas is an inversion of mainstream values regarding childbearing. In the absence of economically viable males, ghetto women view having a child as a conduit to recognition and social prestige among peers — the entry into the "baby club," as he terms it. Although the economic marginalization of black males is a distal cause, the normative adaptation among black ghetto women is a proximate one. Thus, although there may be economic causes and consequences of the high rates of black female-headed households, from a different vantage point there is strong cultural dimension as well.

Our political variable, *black empowerment,* is the proportion of all city council members that are black (mayors included), divided by the proportion of the city's voting age population that is black. This measure has been used in prior research as an indicator of black political power relative to black voting potential within city units (Browning, Marshall & Tabb 1984; LaVeist 1992). If the ratio is less than one, then the African-American community is considered politically underrepresented in the center city, while a ratio above one means it is empowered. We also estimated our models using a measure of the per capita number of black council members per black voting age resident (correlated .478 with black empowerment). However, we do not report this in the tables since the results are nearly identical to those using the first political measure.

We select control variables on both theoretical grounds and because of their use in earlier research. *Black education* is the proportion of black residents 25 years old and older that have completed high school. *Population size* (natural log transformed) reflects variations in the size of cities in our sample, while the proportion of *black males aged 15-34* controls for age structure. Also, given the regional differences in crime rates (Curtis 1975; Huff-Corzine, Corzine & Moore 1986; Liska & Chamlin 1984), we add a control for whether the city is located in the *North* — i.e., outside the southern region. Since disorder and decline may be related to crime (see Skogan 1992), we control for the degree of abandonment and deterioration by including the proportion of all households that are vacant (*vacant households*).

Because multicollinearity is potentially a problem in macrolevel studies, we examine the issue in a number of ways. First, we do not incorporate variables in our models if the correlations with other independent variables are high enough to warrant concern (≥ 0.80). Second, none of the variance inflation factors (VIF) associated with the parameter estimates in our models raised a concern about multicollinearity.[4] However, since multicollinearity cannot be defined solely in terms of the correlations between independent variables, we also probed for multicollinearity while taking the dependent variables (violence rates) into account. Specifically, we found that the parameter estimates for segregation did not change appreciably when removing or including selected predictors. Also, all significant parameter estimates in our models remained virtually the same when we randomly omitted 10% of the cases and reestimated the model (see Maddela 1992).

BLACK ISOLATION VERSUS UNEVENNESS: A SIMULATION

Before analyzing real data for cities in the U.S., we attempt to clarify the distinction between *black isolation from whites* (*P*), the dimension of segregation used in this analysis, and black-white spatial *unevenness* (*D*), the dimension used in previous segregation-crime research. Both *P* and *D* produce the same results in extreme cases of segregation. When blacks and whites in a city live totally apart or, conversely, are completely integrated, *P* and *D* give similar results. But when the degree of black-white residential segregation lies in between these two extremes — and in reality it does — *P* and *D* tell different stories. To clarify this we simulate the geographic distribution of residents in two hypothetical cities depicted in Figure 1.

FIGURE 1: Two Hypothetical Cities Containing Black and White Residents

City 1

Blacks =	5,000	2,500	2,500	0	
Whites =	0	2,500	2,500	5,000	
Percent blacks in each tract =	100%	50%	50%	0%	P = .75, D = 50%

City 2

Blacks =	3,750	3,750	1,250	1,250	
Whites =	1,250	1,250	3,750	3,750	
Percent blacks in each tract =	75%	75%	25%	25%	P = .58, D = 50%

Each city contains just four tracts, and each tract contains some combination of blacks and whites for a total of 5,000 people (we limit our simulation to four tracts for ease of interpretation). The results indicate that both cities have exactly the same level of unevenness in the spatial distribution of blacks versus whites ($D = 50\%$). In either city, exactly half the population of blacks or whites would have to relocate to a different tract in order to achieve an even black-white spatial distribution. However, the degree of black isolation (P) is substantially different between the two cities. Black isolation from whites in city 1 is quite high ($P = .75$), whereas in city 2 it is lower ($P = .58$). This difference is due to the manner in which blacks and whites are distributed across the tracts. City 1 has a tract containing 100% blacks and another containing 0% blacks. In both of these tracts, black isolation is extreme since the chance that blacks will interact with whites is zero. Thus half the tracts in city 1 contain black populations completely isolated from whites, and so overall black isolation in the city is high. On the other hand, *all* the tracts in city 2 contain at least some residents of either race, and thus the degree of black isolation is reduced.

In the real world the two forms of segregation would shade into one another. But this brief simulation shows that segregation in the form of black isolation from whites is distinct, and can vary independently from the more widely used dimension, unevenness. Rather than employing unevenness as the generic form of segregation, we can enhance our understanding about the link between segregation and serious crime by more precisely matching our theoretical arguments to specific distributional forms of segregation.

Results

In Table 1 we provide descriptive statistics for the variables used in our models. Although we analyze only black crime in the multivariate analysis, some descriptive statistics for whites are provided for comparative purposes. Three findings stand out. The most notable feature of Table 1 is the substantial race-gap in the rates of urban violence. The average rate of black homicide and robbery are, respectively, 54.6 and 460.3 per 100,000 population, whereas the corresponding rates for whites are 22.6 and 184.9. These major racial differences in the rates of serious violence are consistent with previous findings and point to the need to racially disaggregate ecological models of violent crime.

A second feature in the findings of Table 1 is the degree of black-white segregation in U.S. cities. In terms of black isolation, the mean probability that a randomly chosen black will be isolated from whites (P) is .56. However, there appears to be high degree of variability in the level of black isolation across our sample of cities (std. dev. = .20). We observed that about one quarter of cities (37) exhibited P values of .75 or greater, suggesting an extreme level of black isolation in a substantial number of urban areas. There also appears to be a considerable amount of unevenness (D) in the residential distribution of urban blacks and whites. The mean index of dissimilarity is 54.3, suggesting that on average over one-half of black residents in center cities would have to relocate in order to create an even residential distribution of blacks and whites.

Table 1 also illustrates the racial differences in urban socioeconomic conditions. Most revealing of the economic and social hardship in urban black communities are the high rates of black poverty. Nearly 30% of black residents (.29) are below the poverty line compared to just over 10% of whites (.11). Indeed, more detailed descriptives of our city data reveal that no city in our sample had a white poverty rate above the mean black poverty rate for all cities. Blacks in our sample also had lower rates of employment compared to whites (.57 vs. .62). Blacks also show four times the white proportion of young, unemployed dropouts (black youth attachment). About 11% of black youths aged 16-19 — or one of every nine — are unemployed, not in school, lack a high school degree, and are not in the armed forces, versus only 4% among whites. Finally, the results reveal the tendency among blacks toward nontraditional family structures. Nearly a third of black households are headed by a female (.31), compared to less than one-tenth among white households (.09).

Overall, these descriptive findings attest to racial differences in the social conditions of blacks versus whites in U.S. cities. In particular, both rates of black poverty and female-headed households are exceedingly high. Likewise, the residents of these communities are restricted to areas with low white populations, increasing the isolation of blacks from whites. The results also confirm the disproportionately high rates of black violence in U.S. cities. The rates of homicide and robbery are several magnitudes higher among blacks than among whites.

TABLE 1: Descriptive Statistics of City-level Structural Variables for Blacks in Major U.S. Cities in 1990

	Mean	Std. Dev.
Black homicide rate (per 100,000)[a]	54.6 (22.8)[b]	157.6 (64.6)
Black robbery rate[a]	460.3 (184.9)	1,650. 2 (805. 2)
D - Uneveness in spatial distribution	54.3	17.1
P - Black-white isolation[a]	.56	.20
Population size	380,275	704,737
Proportion black	.24	.17
Black males aged 15-34	.09	.20
North	.49	.50
Vacant households	.09	.03
Black education	.69 (.78)	.10 (.10)
Black employment	.57 (.62)	.09 (.06)
Black poverty	.29 (.11)	.09 (.04)
Black renters	.64 (.44)	.10 (.09)
Black youth attachment	.11 (.04)	.07 (.03)
Black households, female-headed	.31 (.09)	.06 (.02)
Black political empowerment	3.94	5.17

[a] Original metric shown. Natural log-transformed in the analysis — see text.
[b] Descriptives statistics for whites are provided in parentheses for comparative purposes only.

BLACK-WHITE RESIDENTIAL SEGREGATION AND BLACK VIOLENCE RATES

Table 2 provides the ordinary least squares (OLS) estimates for six models, the first four of which contain a reduced set of predictors. Models 1 and 2 predict black homicide and robbery using only one measure of segregation, unevenness (*D*). Consistent with previous findings, unevenness in the spatial distribution of blacks and whites has a significant positive relationship to both black homicide (.016) and black robbery (.012). As expected, homicide rates are positively related to city size (.966) and to the proportion of black males aged 15-34 (.739). Model 2, which predicts black robbery rates, yields a similar pattern of findings.

The results of models 1 and 2 are consistent with previous research showing that rates of violence are linked to segregation — when segregation is viewed as an *unevenness* (*D*) in the spatial distribution of blacks to whites. However, models 3 and 4 suggest that this relationship may be due to an underlying association between black violence and *black isolation* (*P*) from whites. Model 3 shows that as black isolation increases, so does the black homicide rate (1.145). Likewise in model 4, black isolation is significantly and positively related to black robbery rates (.827). The standardized estimates suggest that in comparison to other variables in the models, black isolation stands out as a major influence of black violence. The effect of black isolation on black homicide is $b = .415$, suggesting that a one-standard-deviation rise in black-white isolation is associated with nearly a one-half standard deviation rise in black homicide, while a similar rise in isolation is associated with more than a one-third standard deviation rise in black robbery (.346). But of particular note is that controlling for black isolation (models 3 and 4) eliminates the effect of unevenness (*D*). Although previous segregation-crime research has focused on unevenness as a predictor of crime, our results imply that a spatial unevenness in the distribution of blacks to whites does not, in itself, engender black violence. The more predictive dimension of segregation appears to be black isolation from whites.

To the extent that black isolation influences black rates of violence is mediated via economic, cultural, and political spheres within black communities, then controlling for these characteristics should decrease the association between black isolation and black violence. Models 5 and 6 predict black homicide and robbery using socioeconomic, cultural, and political variables. *Socioeconomic* variables include black employment, black poverty, and black renters; *cultural* variables include black youth attachment, black female-headed households; and *political* variables include black political empowerment. We observe that the effect of black isolation on serious violence is reduced upon inclusion of these measures. This is particularly evident when comparing the standardized coefficients in models predicting black homicide. In the reduced model (3), the effect of black isolation on homicide is larger ($\beta = .415$) than in the full model ($\beta = .307$). Similarly, the effect of black isolation on robbery ($\beta = .352$) is reduced upon the inclusion of these mediating characteristics ($\beta = .322$).

Of the three sets of variables, those reflecting economic and cultural characteristics stand out as significant. Model 5 of Table 2 shows that black employment is negatively related to the rate of black homicide ($\beta = -.118$). That

TABLE 2: Parameter Estimates Predicting Black Homicide and Robbery in U.S. Cities in 1990, by Segregation Type

	1			2		
	Homicide			Robbery		
	b	β	S.E.	b	β	S.E.
D -> Uneveness	.016*	.175	.008	.012+	.153	.006
P -> Black-white interaction		—			—	
Population size	.966*	.491	.131	.953*	.557	.0113
Black males aged 15-34	.739	.099	.449	.829*	.128	.387
North	-.119	-.039	.129	-.027	-.011	.110
Vacant households	.489	.010	2.071	-3.012+	-.073	1.783
Black education	-.619	-.042	1.050	-1.067	-.082	.893
Black employment		—			—	
Black poverty		—			—	
Black renters		—			—	
Black youth attachment		—			—	
Black households, female headed		—			—	
Black political empowerment						
R^2		.774			.775	

is, as black employment rises, black homicide falls. But both our cultural indicators — black youth attachment and black female-headed households — are also significant predictors. Increases in the proportion of unattached black youth (β = .095) or in the proportion of black households that are headed by a female (β = .123) are accompanied by increases in the rates of black murder. Model 6 also reveals the importance of economics, given that proportion of blacks that are poor (β = .180) or that rent housing (β = .061) are positively related to black robbery, and culture, as suggested by positive relationship between black female-headed households (β = .247) and black robbery.

Yet the effect of black isolation on violence remains significant despite controlling for these features of black urban communities. There may be a number of reasons for this. First, despite efforts to control for economic, cultural and political structures in black urban communities, there are limitations in the kinds of data available to macro researchers of crime (and of other substantive areas) and in the validity of available measures. Second, as suggested in the theoretical discussion, segregation implies not merely the spatial concentration of a group, but also any negative characteristics associated with that group. Thus, a critical feature of social isolation is not just the overall levels of unemployment, youth attachment, etc., but also in how it spatially distributes these characteristics within a city.

Nonetheless, Table 2 offers several important findings. First, black isolation is a robust predictor of black violence, supporting the view that segregation elevates black violence by spatially and socially isolating black communities.

TABLE 2: Parameter Estimates Predicting Black Homicide and Robbery in U.S. Cities in 1990, by Segregation Type (Continued)

	3			4		
	Homicide			Robbery		
	b	β	S.E.	b	β	S.E.
D -> Uneveness	.001	.014	.008	.002	.024	.006
P-> Black-white interaction	1.145*	.415	.827*	.228	.346	.200
Population size	.946*	.481	.936*	.121	.547	.107
Black males aged 15-34	.083	.011	.434	.352	.054	.384
North	.011	.004	.122	.068	.026	.107
Vacant households	-.135	.003	1.918	-3.389+	.082	1.693
Black education	-.664	-.145	.970	-1.040	-.080	.847
Black employment		—			—	
Black poverty		—				
Black renters		—			—	
Black youth attachment		—			—	
Black households, female headed		—			—	
Black political empowerment						
R^2		.809			.799	

Second, previous findings of a link between black-white spatial unevenness (*D*) and black violence appear to be due to an underlying association between black isolation and black violence. Once black isolation is controlled, black violence does not appear to be related to the more typical conceptualization of segregation — spatial unevenness between blacks and whites. Third, the isolation-violence link is partly accounted for by economic and cultural indicators, suggesting that the social isolation of blacks produces economic and normative responses that are conducive to high rates of serious violence.

UNEVENNESS AND BLACK ISOLATION

Although black isolation from whites appears to be a strong predictor of black violence, the degree of independence between the two segregation measures requires closer examination. The correlation between unevenness (*D*) and black isolation (*P*) is substantial (.76, unreported), which raises a concern about the independent effect of black isolation on black violence. In other words, although unevenness and black isolation are potentially independent of one another, to what extent is the effect of black isolation (*P*) on violent crime independent of the underlying level of unevenness (*D*)? If the observed relationship between black isolation (*P*) and crime is significant in cities with low *D* values, as well as in cities with high *D* values, this would strengthen our inference that black isolation has crime-producing effects that are independent of the more commonly used measure of segregation.

TABLE 2: Parameter Estimates Predicting Black Homicide and Robbery in U.S. Cities in 1990, by Segregation Type (Continued)

	5			6		
	Homicide			Robbery		
	b	β	S.E.	b	β	S.E.
D -> Uneveness	.005	.054	.008	.002	.026	.007
P -> Black-white interaction	-.856*	.307	.278	.785*	.322	.234
Population size	.910*	.462	.123	.913*	.533	.104
Black males aged 15-34	.207	.028	.429	.286	.044	.365
North	-.077	-.026	.128	-.070	-.027	.109
Vacant households	2.191	.046	2.112	.305	.007	1.789
Black education	1.367	.092	1.127	.459	.035	.950
Black employment	-2.076+	-.118	1.220	-1.183	-.077	1.184
Black poverty	1.550	.092	1.405	2.607*	.180	1.190
Black renters	-.301	.019	.643	.823+	.061	.482
Black youth attachment	3.536*	.095	1.642	1.431	.045	1.393
Black households, female headed	3.140+	.123	1.826	5.462*	.247	1.613
Black political empowerment	-.011	-.036	.012	.006	.022	.010
R^2		.830			.836	

+ $p < .10$ * $p \leq .05$

To explore these issues in greater detail, we estimate models that predict the effect of black isolation on black homicide and robbery *at two levels of unevenness.* Cities with *D* values of 59 or less are grouped into a "low/medium-D" category, while cities with *D* values of 60 and over are grouped into a "high-D" category. These cutoff points were chosen for two reasons. First, this categorization is consistent with Massey and Denton's (1993) usage where cities with segregation indices of 60 and over are regarded as "hypersegregated." Second, this dichotomization occurs at the near-median point in the sample, resulting in a roughly equal number of cities in each group (low-medium *D*, *N* = 80; High *D*, *N* = 71). Thus, comparisons between models with high-*D* and low/medium-*D* cities are not confounded by major differences in subsample size.

Table 3 presents the results of four models predicting black homicide and robbery for both types of cities. The evidence suggests that the relationship between black isolation and black violence is substantial at both levels of unevenness. For instance, the standardized parameter estimate of black isolation on black homicide in low/medium-*D* cities is .288 (model 1), while in high-*D* cities it is .302 (model 3). This suggests that black isolation has a major influence even in cities with low to moderate levels of black-white unevenness. Indeed, the effect of black isolation on robbery is greater in low/medium-*D* cities (.403, model 2) than in high-*D* cities (.138, model 4).[5] Overall, the fact that black isolation was a significant predictor in both medium/low-*D* and high-*D* cities, strengthens our confidence regarding an independent effect of black isolation on

the city rates of black homicide and robbery. Segregation, in the form of a physical and social isolation of blacks from whites in urban areas is associated with high rates of violence in urban black communities.

Summary and Discussion

We extend the research on segregation and crime by addressing three key limitations of extant research in this area. First, prior conceptualizations about segregation and crime are drawn from the widely known literature regarding inequality and crime, obscuring the potentially more relevant issue of black social isolation. Second, previous research relies on indicators that conceptualize residential segregation as a spatial unevenness (*D*) in the distribution of blacks to whites, neglecting the theoretically critical dimension of black isolation (*P*) from whites. Third, the bulk of previous studies have not examined how segregation directly effects black violence, despite the fact that both segregation and its consequences are problems faced mainly by urban black communities. Building on recent research on the rise of the urban underclass, we examine whether reductions in the potential for race contact, black isolation, leads to high rates of urban black homicide and robbery.

Our results reveal that segregation is a major predictor of the rates of homicide and robbery among blacks. Consistent with prior research, spatial unevenness (*D*) is positively associated with serious black violence at the city level. However, when black isolation (*P*) is incorporated into our models of serious black crime, the effect of unevenness becomes negligible. Instead, the results reveal that in cities when blacks are isolated from whites (i.e., black isolation), the rates of serious black violence are exceedingly high. Several potential mediators of this effect include the black rates of unemployment, poverty, youth institutional attachment, and female-headed households. However, since these factors accounted for only about 25% of the empirical association between black isolation and black homicide, future research needs to identify additional factors that mediate the link between isolation and violence. Our findings also show that the effect of black isolation on serious violence is substantial even in cities in which the underlying level of unevenness (*D*) is medium/low.

We derive a number of implications from these findings. First, the results support the view that black social isolation has profoundly negative consequences. Secluded in their own communities, many urban blacks lack the sustained contact with mainstream institutions necessary for upward social and economic mobility. Black isolation may lead to weakened labor force attachment and thereby deprive residents of the main conduit by which adults prove their worth in the larger society. In extreme cases it can create an environment which fosters a self-image based on violence and reinforce a variety of behaviors that increase the social disability of blacks. Black isolation can also divest black urban communities of the advantages of pluralist politics and makes them vulnerable to economic downturns and government cutbacks. The combination of these and other negative consequences of black isolation, exacerbates the disinvestment

TABLE 3: Parameter Estimates Predicting Black Homicide and Robbery at Two Levels of Unevenness (D) in Spatial Distribution

	1			2		
	Low-D (0-59)					
	Homicide			Robbery		
	b	β	S.E.	b	β	S.E.
P -> Black-white interaction	.990+	.288	.554	1.174*	.403	.436
Population size	.289	.143	.237	.590*	.343	.190
Black males aged 15-34	14.778*	.396	4.989	8.170*	.259	3.995
North	-.099	-.041	.212	.152	.074	.169
Vacant households	4.758	.115	2.997	-.273	-.008	2.392
Black education	1.414	.117	1.555	2.941*	.286	1.242
Black employment	-3.256+	-.243	1.908	-2.098	-.184	1.638
Black poverty	1.689	.126	1.956	2.355+	.208	1.361
Black renters	.556	.045	.929	.108	.011	.722
Black youth attachment	4.558*	.174	2.120	1.283	.058	1.698
Black households, female headed	2.448	6.481*	.129	2.597	.406	2.072
Black political empowerment	-.073	.079	-.043	.126	.054	.101
R^2		.715			.742	

+ $p < .10$ * $p < 0.05$

and decline of many black communities in major cities of the U.S. and contributes to the rise of black predatory crime.

Second, researchers should recognize that blacks and other minorities can be segregated in a variety of ways. Broadening our view of residential segregation will lead us to consider alternative — and perhaps more appropriate — causal mechanisms by which segregation leads to crime. Of particular concern, is the heavy reliance on unevenness as an indicator of black-white residential segregation and as a predictor of violence. In light of our findings, this practice appears to be unwarranted. Indeed, we suspect that previous evidence of a significant relationship between violence and segregation — as measured by unevenness (*D*) — actually reflects the underlying relationship between black violence and black isolation. To understand how the spatial distribution of blacks relative to whites generates violence and other indicators of social malaise, sociologists need to examine the theoretical mechanisms implied in each of the various forms of residential segregation. For instance, another distributional form of segregation identified by Massey and Denton (1988b) is "centralization." This refers to the degree to which minority groups reside close to the center of a city. A research tradition descended from the Chicago school (cf. McKenzie 1933; Park & Burgess 1925; Wirth 1928) documents how the abundance of low-skill jobs contributed to the social mobility of successive waves of immigrants. But recently there has been a growing concern about the

TABLE 3: Parameter Estimates Predicting Black Homicide and Robbery at Two Levels of Unevenness (D) in Spatial Distribution (Continued)

	3			4		
	High-D (60+)					
	Homicide			Robbery		
	b	β	S.E.	b	β	S.E.
P -> Black-white interaction	.861*	.302	.287	.366+	.138	.216
Population size	1.206*	.732	.136	1.169*	.761	.128
Black males aged 15-34	-.110	-.022	.395	.750	.016	.373
North	.108	.039	.170	.037	.014	.160
Vacant households	-.417	.009	2.644	3.021	.070	2.499
Black education	.223	.010	1.603	-2.413	-.114	1.515
Black employment	-1.799	-.073	2.468	.059	.003	2.332
Black poverty	1.800	.079	2.366	1.206	(.057)	2.236
Black renters	2.413*	.148	.850	1.675*	.110+	.803
Black youth attachment	.976	.022	2.591	-.573	-.014	2.448
Black households, female headed	-.074	-.002	2.684	2.891	.078	2.537
Black political empowerment	.064	.038	.079	.150*	.094	.075
R^2		.892			.889	

+ $p < .10$ * $p < .05$

structural disadvantages inner-city life, which include a shrinking low-skill job-base, the depletion of the city's tax base as wealthy residents move to the suburbs, and a host of other factors that reduce the economic and social opportunities of inner-city residents (Blackley 1990; Cain & Finnie 1990; Hawley 1950; Kasarda 1990). In a broader sense, the structural disadvantages specific to inner-city life may contribute to the economic decline of minority communities in center cities and erode their capacity to restrain violent behavior.

Finally, we suggest several avenues for future research that build on the present analysis. First, although the objective of the present analysis was to examine whether a cross-sectional relationship exists between social isolation and violence, we would like to see this work extended using a longitudinal design. Massey and Denton (1993) demonstrate that black isolation increased steadily after the turn of the century. Although this implies that such a study would use data for many decades, it is likely to be truncated given the poor and incomplete UCR data that exist prior to 1960. Second, Wilson (1993) points out that social isolation results not only from a lack of contact between the races, but between the classes as well. Particularly informative would be studies examining whether the social isolation of low-income blacks from medium/high-income blacks in urban areas has a crime-producing effect in black communities. Third, while previous studies examine whether the spatial distribution of *people* impacts on serious crime, an important though neglected question concerns how the spatial distribution of socioeconomic *characteristics* influences crime. Underlying the

relationship between segregation and crime is not merely the spatial isolation of blacks, but the geographic concentration of negative characteristics that are collinear with the urban black experience, such as poverty, joblessness, low job-skills, low education, welfare, teen pregnancy, and family disruption. Such studies would refine our understanding about how the spatial segregation of U.S. blacks, unique in its degree and persistence, leads to lethal violence and other predatory behavior.

Notes

1. Massey and Denton (1988) identify five distinct forms of segregation. (1) *Unevenness,* which refers to the distribution of minority and majority groups relative to one another across spatial units of a city. (2) *Isolation,* a concept referring to the degree to which minority group members come into contact with only other minority group members. (3) *Centralization* measures the degree to which a minority group is spatially located in the center of an urban area. (4) *Concentration* reflects the amount of physical space a minority group occupies in an urban area. A highly concentrated group resides in a relatively small area. (5) *Clustering* refers to a residential pattern in which spatial units occupied by minority groups are contiguous to each other. High clustering resembles an enclave, while low clustering resembles a checkerboard pattern.

2. In Peterson and Krivo's (1993) important analysis of segregation and crime, social isolation is noted as a possible mechanism by which segregation increases crime (p. 1005). However, the authors measure segregation using both the Index of Dissimilarity and ETA. But since Massey and Denton (1988) show in unambiguous terms that both of these measures reflect the same thing — unevenness in spatial distribution, a direct measure of black-white social isolation is not used. Potter (1991) did use a black-white isolation measure, but his analysis cannot readily be grouped with the other segregation-crime studies for several reasons. First, it is a demographic analysis (published in *Demography*) that focuses mainly on black-white differences in life expectancies. Second, the empirical and theoretical relationship between segregation and crime is not explored. Third, standard independent and control variables in criminology are not included in the analysis. Fourth, the impact of segregation on *black* crime is not examined.

3. While the diagnostic statistics justified this transformation, we also estimated our models using crime rates that were not log-transformed. The results did not differ substantially. Sampson (1987) and Shihadeh and Steffensmeier (1994) also multiplied the arrest rates by the ratio of offense to arrest in each city in order to account for possible jurisdictional biases in the probability of arrest across cities. However, we conclude that such an adjustment is unnecessary and does not meaningfully alter the results. We are not surprised that adjusting the data makes little differences since the correlation between adjusted and unadjusted violence rates is over .90.

4. None of the variance inflation factors (VIF) associated with the parameter estimates exceeded a value of 4. The values ranged from 1.16 to 3.24.

5. To examine whether these results are sensitive to the specific dichotomy chosen, we estimated models at a variety of cutoff points of D. The extended analysis showed that the association between P and crime was significant when we varied the cutoff points of D values from 48 to 68. Samples of cities with D values lower than 48 or higher than 68 were excessively small, which seriously reduced our degrees of freedom (e.g., $D < 48$, $N \leq 37$; $D > 68$, $N = 28$). Also, although the results in Table 3 suggest that the association between P and crime is greater in low/medium cities than in high-D cities, further analysis shows that this difference is not significant. We estimated several models that contained a multiplicative interaction term between D and P. In some models we maintained D as a continuous variable, both in the main effect and interaction term, while in others we dichotomized D into the two categories listed in Table 3. In each case the interaction term was not significant, suggesting that the effect of P on crime does not change at either level of D.

References

Allan, Emilie Anderson, and Darrell J. Steffensmeier. 1987. "Youth, Underemployment and Property Crime: Differential Effects of Job Availability and Job Quality on Juvenile and Young Adult Arrest Rates." *American Sociological Review* 54:107-23.

Balkwell, James W. 1993. "Metropolitan Structure and Violent Crime: A Further Examination of the Blau and Blau Relative Deprivation Thesis." Paper presented at the annual meeting of the American Society for Criminology, Denver Colorado.

Banfield, Edward C. 1974. *The Unheavenly City Revisited*. Little, Brown.

Blackley, Paul R. 1990. "Spatial Mismatch in Urban Labor Markets: Evidence from Large U.S. Metropolitan Areas." *Social Science Quarterly* 71:39-52.

Black, Donald, and Albert J. Reiss. 1970. "Police Control of Juveniles." *American Sociological Review* 35:63-77.

Blau, Judith, and Peter M. Blau. 1982. "The Cost of Inequality: Metropolitan Structure and Violent Crime. *American Sociological Review* 47:114-29.

Bourdieu, Pierre. 1986. "The Forms of Capital." Pp. 26-65 in *Handbook of Theory and Research for the Sociology of Education*, edited by John G. Richardson. Greenwood Press.

Bradford, C., and L. Rubinowitz. 1975. "The Urban-Suburban Investment-Disinvestment Process: Consequences for Older Neighborhoods. *Annals of the American Academy of Political and Social Science* 422:77-86.

Browning, Rufus P., Dale R. Marshall, and David H. Tabb. 1984. *Protest Is Not Enough*. University of California Press.

Bursick, Robert J., Jr., and Harold G. Grasmick. 1993. *Neighborhoods and Crime: The Dimensions of Effective Community Control*. Lexington Books.

Cain, Glen G., and Ross E. Finnie. 1990. "The Black-White Difference in Youth Employment: Evidence for Demand-Side Factors." *Journal of Labor Economics* 43:S364-S395.

Carroll, Leo, and Pamela Irving Jackson. 1983. "Inequality, Opportunity, and Crime Rates in Central Cities." *Criminology* 21:178-94.

Carpenter, Allan. 1992. *Facts about the Cities*. H.W. Wilson.

Cherlin, Andrew J. 1992. *Marriage, Divorce, Remarriage*. Harvard University Press.

Coleman, James S. 1988. "Social Capital and the Creation of Human Capital." *American Journal of Sociology* 94:S95-S120.

Crain, Robert L., and Rita E. Mahard. 1978. "School Racial Composition and Black College Attendance and Achievement Test Performance." *Sociology of Education* 52:81-101.

Curtis, Lynn. 1975. *Violence, Race, and Culture*. D.C. Heath & Company.

Dahl, Robert A. 1967. *Pluralist Democracy in the United States: Conflict and Consent*. Rand McNally.

Devine, Joel A., Joseph F. Sheley, and M. Dwayne Smith. 1988. "Macroeconomic and Social-Control Policy Influences in Crime Rates, 1984-1985." *American Sociological Review* 53:407-21.

Edsall, Thomas B., and Mary D. Edsall. 1991. *Chain Reaction: The Impact of Race, Rights, and Taxes on American Politics*. Norton.

Ensminger, M.E., S.G. Kellam, and B. Rubin. 1983. "School and Family Origins of Delinquency: Comparisons by Sex." In *Antecedents of Aggression and Antisocial Behavior*, Edited by K.T. Van Dusen and S.A. Mednick. Kluwer-Nijhoff.

Farley, Reynolds, and William H. Frey. 1994. "Changes in the Segregation of Whites From Blacks During the 1980's: Small Steps towards a More Integrated Society." *American Sociological Review* 59:23-45.

Farley, Reynolds, Howard Schuman, Suzanne Bianchi, Diane Colasanto and Shirley Hatchett. 1978. "Chocolate City, Vanilla Suburbs: Will the Trend Towards Racially Separate Communities Continue?" *Social Science Research* 7:319-44.

Farley, Reynolds, Charlotte Steeh, Tara Jackson, Maria Krysan, and Keith Reeves. 1994. "The Causes of Continued Racial Residential Segregation: Chocolate City, Vanilla Suburbs Revisited." *Journal of Housing Research* 4:1-38.

Fordham, Signithia, and John Ogbu. 1986. "Black Student's School Success: Coping with the 'Burden of Acting White'." *The Urban Review* 18:176-206.

Gastil, Raymond. 197. "Homicide and the Regional Culture of Violence." *American Sociological Review*, 36:412-27.

Gottfredson, Michael, and Don Gottfredson. 1980. *Decision making in Criminal Justice*. Ballinger.

Gove, Walter, Michael Hughes, and Michael Geerken. 1985. "Are Uniform Crime Reports a Valid Indicator of the Index Crimes? An Affirmative Answer with Minor Qualifications." *Criminology* 23:451-502.

Grimshaw, William J. 1980. *Black Politics in Chicago: The Quest for Leadership, 1939-1979*, Loyola University Press.

Gutman, Herbert B. 1976. *The Black Family in Slavery and Freedom, 1750-1925*. Pantheon.

Hagan, John. 1994. *Crime and Disrepute*. Pine Forge Press.

Harer, Miles D., and Darrell Steffensmeier. 1992. "The Different Effects of Economic Inequality on Black and White Rates of Violence." *Social Forces* 70:1035-54.

Hawley, Amos H. 1950. *Human Ecology: A Theory of Community Structure*. Ronald Press.

Hindelang, Michael J. 1978. "Race and Involvement in Common Law Personal Crimes." *American Sociological Review* 43:93-109.

Huff-Corzine, Lin, Jay Corzine, and David C. Moore. 1986. "Southern Exposure: Deciphering the South's Influence on Homicide Rates." *Social Forces* 64:906-24.

Hunter, Albert J. 1985. "Private, Parochial and Public School Orders: The Problem of Crime and Incivility in Urban Communities." Pp. 203-42 in *The Challenge of Social Control: Citizenship and Institution Building in Modern Society*, edited by Gerald D. Suttles and Mayer N. Zald. Ablex Publishing

Joint Center for Political Studies. 1989. *Black Elected Officials: A National Roster*. University Press of America, Washington D.C.

____. 1990. "Structural Factors Affecting the Location and Timing of the Urban Underclass Growth." *Urban Geography* 11:234-64.

Kasarda, John D., and Morris Janowitz. 1974. "Community Attachment and Mass Society." *American Sociological Review* 47:427-33.

Katznelson, Ira. 1976. *Black Men, White Cities: Race, Politics, and Migration in the United States, 1900-1930, and Britain 1948-68*. University of Chicago Press.

Kemp, Kathleen A. 1986. "Race, Ethnicity, Class, and Urban Spatial Conflict: Chicago as a Crucial Test Case", *Urban Studies* 23:197-208.

Labov, William. 1980. "The Logic of Nonstandard English." In *Locating Language in Space and Time*, edited by William Labov. Academic Press.

LaFree, Gary, Kriss A. Drass, and Patrick O'Day. 1992. "Race and Crime in Postwar America: Determinants of African-American and White Rates, 1957-1988." *Criminology* 30:157-88.

LaVeist, Thomas A. 1992. "The Political Empowerment and Health Status of African-Americans: Mapping a New Territory." *American Journal of Sociology* 97:1080-95.

Lichter, Daniel T. 1988. "Racial Differences in Unemployment in American Cities." *American Journal of Sociology* 93:771-92.

Lieberson, Stanley. 1981. *An Asymmetrical Approach to Segregation."* Pp. 61-82 in *Ethnic Segregation in Cities*, edited by Ceri Peach, Vaughn Robinson, and Susan Smith. Croom Helm.

Liska, Allen E., and Mitchell B. Chamlin. 1984. "Social Structure and Crime Control among Macrosocial Units." *American Sociological Review* 90:383-95.

Logan, John R., and Steven F. Messner. 1987. "Racial Residential Segregation and Suburban Violent Crime." *Social Science Quarterly* 68:510-27.

Maddela, G.S. 1992. *Introduction to Econometrics*. 2d ed. Macmillan.

Massey, Douglas S. 1990. "American Apartheid: Segregation and the Making of the Underclass." *American Journal of Sociology* 96:329-57.

Massey, Douglas S., and Nancy A. Denton. 1988. "Suburbanization and Segregation in U.S. Metropolitan Areas." *American Journal of Sociology* 94:592-626.

____. 1993. *American Apartheid: Segregation and the Making of the Underclass*. Harvard University Press.

Massey, Douglas S., and Brendan P. Mullan. 1984. "Processes of Hispanic and Black Spatial Assimilation." *American Journal of Sociology* 89:836-73.

Massey, Douglas S., Andrew B. Gross, Kumiko Shibuya. 1994. "Migration, Segregation and Geographic Concentration of Poeverty." *American Sociological Review* 59:425-445.

Matsueda, Ross, and Karen Heimer. 1987. "Race, Family Structure, and Delinquency: A Test of Differential Association and Social Control Theories." *American Sociological Review* 52:826-40.

McKenzie, Roderick D. 1933. *The Metropolitan Community*. McGraw-Hill Book Company.

McLanahan, Sarah, and Karen Booth. 1989. "Mother-Only Families: Próblems, Prospects, and Politics." *Journal of Marriage and the Family* 51:557-580.

Messner, Steven F., and Scott J. South. 1986. "Economic Deprivation, Opportunity Structure, and Robbery Victimization: Intra- and Interracial Patterns." *Social Forces* 64:975-91.

Morgan, S. Philip, Antonio McDaniel, Andrew T. Miller, and Samuel H. Preston. 1990. "Racial Differences in Household and Family Structure at the Turn of the Century," paper presented at the Albany Conference on Demographic Perspectives on the American Family: Patterns and Prospects, SUNY.

Park, Robert E., and Ernest W. Burgess. 1925. *The City*. University of Chicago Press.

Patterson, Britt E. 1991. "Poverty, Income Inequality, and Community Crime Rates." *Criminology* 29:755-76.

Pedder, Sophie. 1991. "Social Isolation and the Labor Market: Black Americans in Chicago", paper presented at the Chicago Urban Poverty and Family Life Conference, Chicago, 10/91.

Petersilia, Joan. 1985. "Racial Disparities in the Criminal Justice System: A Summary." *Crime and Delinquency* 31:15-34.

Peterson, Ruth D., and Lauren J. Krivo. 1993. "Racial Segregation and Urban Black Homicide" *Social Forces* 71:1001-26.

Philpot, Thomas L. 1978. *The Slum and the Ghetto: Neighborhood Deterioration and Middle-Class Reform, Chicago, 1800-1930*. Oxford University Press.

Pinderhughes, Dianne M. 1987. *Race and Ethnicity in Chicago Politics*. University of Illinois Press.

Potter, Lloyd B. 1991. "Socioeconomic Determinants of White and Black Males' Life Expectancy Differentials, 1980." *Demography* 28:303-21.

Rosenfeld, Richard. 1986. "Urban Crime Rates: Effects of Inequality, Welfare Dependency, Region, and Race." Pp. 116-30 in *The Social Ecology of Crime*, edited by James M. Byrne and Robert J. Sampson. Springer-Verlag.

Ross, Heather, and Isabel Sawhill. 1975. *Time of Transition: The Growth of Families Headed by Women*. Washington, D.C.: Urban Institute.

____. 1985. "Race and Criminal Violence: A Demographically Disaggregated Analysis of Urban Homicide. " *Crime and Delinquency* 31:47-82.

____. 1987. "Urban Black Violence: The Effect of Male Joblessness and Family Disruption." *American Journal of Sociology* 93:348-82.

Shaw, Clifford R., and Henry D. McKay. 1931. *Social Factors in Juvenile Delinquency. Vol. 2 of Report on the Causes of Crime.* National Commision of Law Observance and Enforcement. Government Printing Office.

Shihadeh, Edward S., and Graham Ousey. 1996. "Metropolitan Expansion and Black Social Dislocation: The Link between Suburbanization and Center-City Crime." *Social Forces.* Forthcoming.

Shihadeh, Edward S., and Darrell J. Steffensmeier. 1994. "Economic Inequality, Family Disruption and Urban Black Violence: Cities as Units of Stratification and Social Control." *Social Forces* 73:729-51.

Shryock, Henry S., and Jacob S. Siegal. 1976. *The Methods and Materials of Demography*. Academic Press.

Skogan, Wesley G. 1992. *Disorder and Decline: Crime and the Spiral of Decay in American Neighborhoods.* University of California Press

South, Scott J., and Richard B. Felson. 1990. "The Racial Patterning of Rape." *Social Forces* 69:71-93.

Sudarkasa, Niara. 1981. "Interpreting the African Heritage in Afro-American Family Organization" Pp. 37-53 in *Black Families,* edited by Harriette Pipes McAdoo. Sage Publications.

U.S. Department of Justice. 1988. *Uniform Crime Reports*. Federal Bureau of Investigation.

Wallace, Deborah. 1990 "Roots of Increased Health Care Inequality in New York." *Social Science and Medicine* 31:1219-27.

Wetzel, James R., 1995. "Labor Force, Unemployment, and Earnings." In *State of the Union: America in the 1990s,* edited by Reynolds Farley, Russell Sage Foundation, New York.

Wilson, William Julius. 1984. "The Urban Underclass." In *Minority Report,* edited by Leslie W. Dunbar. Pantheon.

____. 1987. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. University of Chicago Press.

____. 1993. "The Underclass: Issues, Perspectives, and Public Policy." In *The Ghetto Underclass,* edited by William Julius Wilson. Sage Publications.

Wirth, Louis. 1928. *The Ghetto*. University of Chicago Press.

Yancey, William L., Eugene P Erickson, and George H. Leaon, 1985. "The Structure of Pluralism: 'We're All Italian Around Here, Aren't We Mrs. O'Brien?'" *Ethnic and Racial Studies* 8:94-116.

Perceived Risk versus Fear of Crime: Empirical Evidence of Conceptually Distinct Reactions in Survey Data*

PAMELA WILCOX ROUNTREE, *University of Kentucky*
KENNETH C. LAND, *Duke University*

Abstract

The conceptualization and measurement of fear of crime have received considerable attention in the research literature. Nevertheless, most sample surveys use indicators that only tap a general, cognitive assessment of safety — assumed to represent fear of violence. This article examines whether there are conceptually and empirically distinct dimensions of fear of crime by comparing the micro- and macrolevel antecedents of such a general, cognitive "fear" — which we call "perceived risk" — with those for a more emotionally based, burglary-specific fear. Hierarchical logistic regression models for both types of fear are presented. Some similarities were found between fear and risk in terms of various predictors, but results generally provide further empirical evidence that the two constructs are quite distinct. At the individual-level, the effect of gender is different across models, and routine-activities variables are somewhat better predictors of burglary-specific fear in comparison to risk perception. At the contextual level, neighborhood integration serves to diminish respondents' perceptions of neighborhood danger, yet this variable is positively related to burglary-specific fear.

Previous research on fear of crime suggests that there is ambiguity in defining and measuring fear. Different studies of fear use different operational measures, illustrating a lack of consensus on what "fear of crime" actually means (see Ferraro & LaGrange 1987 for a review of fear of crime studies and their respective measures; also see Bursik & Grasmick 1993). Some measures stem from survey questions such as "How safe is your neighborhood from crime?" and "How safe do you feel walking alone in your neighborhood at night/during the day?" In these questions, respondents are asked to make judgments about risk while in questions such as "How afraid are you of becoming a victim of rape/robbery/murder, etc.?" they are asked to make more personal statements of emotionally based concern about crime. Yet, when scholars discuss fear of crime using any of these different types of measures,

** We would like to thank Terance D. Miethe for providing the survey data analyzed herein. This research was supported in part by National Science Foundation Grant SBR-9310193.* *Direct correspondence to Pamela Wilcox Rountree, Department of Sociology, University of Kentucky, Lexington, KY, 40506-0027.*

 Social Forces, June 1996, 74(4):1353-1376

often no attempt is made to specify the type of fear or the type of crime under question. "Fear of crime," then, ranges from general, cognitive perceptions of safety in one's neighborhood to affective, personal, emotional reactions to the possibility of being victimized by a specific type of crime (Ferraro & LaGrange 1987).

Given the range of possibilities, concerning fear of crime, it is quite conceivable that the causes of this fear vary, depending upon the types of perceptions or reactions under consideration. Nonetheless, very few studies have attempted to address these potential distinctions. The review of the literature on fear of crime by Ferraro and LaGrange (1987) indicates that a large percentage of studies use only one indicator of fear — usually a more general, cerebral assessment of community safety for tapping fear rather than a personal reaction to a specific type of victimization. Indeed, Ferraro and LaGrange (1987), using a definition of fear of crime that "refers to the negative emotional reactions generated by crime or symbols associated with crime" (73), claim that the often-used general, cognitive measures refer to "judgments of risk," not actual fear. Thus, recent attention to the measurement of fear has underscored a conceptual distinction between commonly used general, cognitive "fear" — or risk perception — and crime-specific, emotionally based fear.

However, the literature is limited in terms of explicit and comprehensive examination of potential distinctions regarding covariates of different measures of the fear of crime. The present study, while assuming the Ferraro and LaGrange (1987) definition of fear — "the negative emotional reaction generated by crime or symbols of crime" (73) — extends previous work by integrating micro- and macropredictors in examining the differences between personal, burglary-specific fear of crime and a measure of risk perception resembling a conceptualization and operationalization of fear employed in most extant research.

Previous Research: Correlates of Fear of Crime

As mentioned above, most research on the fear of crime focuses on a cognitive or judgment-based, general concern that may better be termed *risk perception*. We used survey items such as "Do you think that people in this neighborhood are safe inside their homes at night/during the daytime?" or the previously mentioned, "How safe is your neighborhood from crime?" are commonly used. These items elicit general assessments of safety from the respondents rather than personal, more emotional experiences (Ferraro & LaGrange 1987). In support of the purported conceptual distinction of this type of measure from actual fear, others have speculated that fear of crime, when measured in this general sense, may refer to perceptions of urban disorder rather than actual fear of crime per se (e.g., Bursik & Grasmick 1993; Garofalo & Laub 1979; Wilson & Kelling 1982).

Many of the earlier studies of this judgment-based, general type of fear (or risk perception) focused on sociodemographic correlates, finding, for example, that fear was directly related to age and gender. Typically, these studies found that women are more fearful and that fear increases with age (e.g., Baumer

1979; Clarke & Lewis 1982; Garofalo 1979). These sociodemographic correlates are presumably related to physical or social vulnerability. More recent research on the sociodemographic correlates has focused on, for example, interactions between age, race, and sex, indicating that the effects of sociodemographic characteristics on fear may be other than simply additive (Baumer 1985; Ortega & Myles 1987).

Research on the relationship between fear of crime and information about crime — through direct or indirect victimization, or through media accounts — has also been prevalent. Despite the fact that prior victimization is generally considered a relatively weak predictor of fear, work by Skogan and Maxfield (1981) indicated that previous victims are more fearful than their nonvictimized counterparts. Skogan and Maxfield also found significant effects on fear for indirect victimization — referring to knowledge of crime or victimization experiences of family, friends, and neighbors via interpersonal communication. Other research indicates that initial media reports of serious local crimes can increase levels of fear (e.g., Liska & Baccaglini 1990). Further individual-level variables found to affect fear in the literature include social-psychological characteristics such as locus of control and self-image (van der Wurff, van Staalduinen & Stringer 1989) and lifestyle or routine-activity attributes such as exposure (Stafford & Galle 1984).

Extant research on the fear of crime has also emphasized the contextual (city- or neighborhood-level) factors involved in explaining fear. Macrolevel studies indicate that structural features such as physical and social incivility, poor social integration, segregation, population size, and community-level crime increase levels of fear among residents (Lewis & Maxfield 1980; Lewis & Salem 1986; Liska, Lawrence & Sanchirico 1982; Liska & Warner 1991; Skogan & Maxfield 1981; Taylor, Gottfredson & Brower 1984; Taylor & Hale 1986).

Risk Perception versus Offense-Specific Fear of Crime

As alluded to earlier, the extant research on the fear of crime has generally ignored the potentially important distinctions between types or dimensions of fear of crime. While most researchers recognize that fear of crime can have a multitude of meanings, the studies that have been done focus on a general, judgment-based measure for fear and assume that fear of violent crime is being tapped. Research efforts by Warr and Stafford (1983), Warr (1987), LaGrange, Ferraro, and Supancic (1992), and Ferraro (1995) represent the few exceptions to this trend and support the notion of a conceptual distinction between judgments of risk and actual fear. Warr and colleagues have examined the proximate causes of fear for up to sixteen different offense categories and have found that, although risk perception ("how likely is ____ to happen to you in the next year?") is often an important predictor of fear ("how afraid are you of becoming a victim of ____?"), it is not perfectly correlated with fear, since fear also depends upon individual-level perceived seriousness of the offense in question and an individual's risk sensitivity (Warr 1987; Warr & Stafford 1983). In a study of the contextual predictors of fear, LaGrange, Ferraro, and Supancic (1992) found that neighborhood incivilities — commonly found predictors of

"fear" in the extant research reviewed above — were more strongly related to perceived risk than they were related to actually being afraid. LaGrange and associates (1992) also found perceived risk to be an important predictor of actual fear, thereby further supporting the need to disambiguate between fear and judgments of risk. That is, assuming that measures of judgments of risk also measure fear of crime "is both invalid and obscures the processes that generate these perceptions" (Ferraro & LaGrange 1987:73).

In all these studies, then, perceived risk is shown to be one of several important components in explaining fear. Though such findings support the notion of a conceptual distinction between perceived risk and actual fear as posited by Ferraro and LaGrange (1987), none of the studies gives explicit attention to this matter in any comprehensive empirical way. A recent study by Ferraro (1995) deals with this issue somewhat more explicitly and finds evidence of conceptual distinction, yet uses a limited array of contextual and individual-level variables. Thus, the present study tests the ideas of Ferraro and LaGrange (1987) and extends the empirical evidence of others (Warr & Stafford 1983; Warr 1987; LaGrange, Ferraro & Supancic 1992; Ferraro 1995) while examining differences in empirical results between judgment-based risk perception and more personal, emotionally based burglary-specific fear. In particular, we seek to uncover and understand distinctions in the significance of individual-level predictors as well as in direct and moderating effects of social context regarding risk perception as opposed to a specific concern for burglary victimization.

Methods

In order to address these potential distinctions, we employ a hierarchical logistic regression modeling procedure using the ML3E software (Prosser, Rasbash & Goldstein 1991a, 1991b). The hierarchical modeling approach, growing out of Bayesian estimation in statistics (Lindley & Smith 1972), sociological contextual-effects research (Mason, Wong & Entwisle 1983), and educational research on growth curves and classroom and school effects (e.g., Goldstein 1987; Raudenbush 1984; Strenio, Weisberg & Bryk 1983), has recently been applied to criminological multilevel research on the antecedents and effects of victimization (Rountree & Land 1996; Rountree, Land & Miethe 1994).

While distinctions between risk perception and burglary-specific fear in terms of multilevel explanatory variables could be studied through conventional linear or logistic models, hierarchical models more appropriately account for the implicit hierarchy and structure of data. Simply including contextual variables in conventional linear or logistic regression models along with individual-level variables does not account sufficiently for the idea that individuals fall into many socially meaningful groups or categories and that individuals clustered within the same groups are often likely to be more similar regarding certain social characteristics than are individuals grouped differently. For instance, because individuals are distributed nonrandomly across neighborhoods, an assumption of independent error terms is inappropriate when the survey design incorporates individuals clustered within various neighborhoods (Hox & Kreft

1994; Ringdal 1992). Hierarchical linear or logistic regression models account for the implicit hierarchy of data by employing submodels — each of the levels in the structure or hierarchy is represented with its own submodel. These submodels, along with nested error terms, can account for effects and sources of variation at each of the levels of analysis represented in the data (see Bryk & Raudenbush 1992 for a general exposition).

Data and Variables

SAMPLE DESIGN

The primary data source for the analyses presented here is victimization survey data as part of a larger crime research project for Seattle, Washington. The survey data were collected in 1990 on 5,302 individuals clustered within 300 city-block pairs (or neighborhoods) and 100 of the city's stable 121 census tracts (Miethe & McDowall 1993; Miethe & Meier 1994). We deleted cases with missing data, leaving 5,090 individuals for analysis.[1] Three pairs of city blocks were selected from each of the census tracts. One block in each pair contained a street address at which a burglary had been reported in 1989; the other block in each pair bordered the block with the burglary. A sample was then drawn from residents of the three city-block pairs within each tract.

While the city-block pair was the contextual unit of interest in the original study, here we are interested in both the city-block pair, or *neighborhood*, and the *census tract*. The fact that one of our contextual variables of interest — official neighborhood burglary rates — was unavailable at units smaller than the census tract primarily guided our decision to include the tract as a contextual unit. Due to the clustering at the neighborhood level in the sample design, however, ignoring neighborhoods as a contextual unit would undermine results if, as is highly possible, households in the same neighborhood within a given tract are more similar to one another than they are to other households in the same tract yet not within the same immediate neighborhood.

Within each neighborhood, an average of 18 households were sampled; within each tract, an average of 50 housing units were sampled. A reverse telephone directory was the sampling frame. The resulting sample appears to underrepresent young adults, low-income families, and movers (transient households); home owners and college-educated respondents are overrepresented (Miethe & Meier 1994). Finally, because of the way in which neighborhoods were selected within tracts (on the basis of burglary reports), the sample design also disproportionately stratifies on the basis of burglary victimization.

DEPENDENT VARIABLES

Both perceived crime risk and burglary-specific fear serve as dependent variables for our analyses. *Burglary-specific fear* is indicated by whether or not the respondent worries at least once a week about his/her home being burgled. *Perception of crime/victimization risk,* on the other hand, is indicated by whether

or not the respondent perceives his/her neighborhood to be either somewhat unsafe or very unsafe from crime.[2]

Both of these measures refer to perceptions or emotions specific to the time of the survey. Nearly 39% of sampled respondents were fearful of burglary specifically, while just over 24% indicated a feeling of being unsafe in their neighborhood (see Table 1).

INDIVIDUAL-LEVEL EXPLANATORY VARIABLES

Both individual-level and aggregate-level, contextual indicators are used as independent variables in the analyses. First, individual-level independent variables include sociodemographic characteristics, lifestyle or routine-activity attributes, including factors related to the ecology of crime, and information about crime stemming from previous personal victimization experiences.

Previous burglary victimization is indicated by whether or not the respondent experienced an attempted or completed break-in of his/her current home within the two years *preceding* the survey (in contrast to the contemporaneous nature of the fear measures). Thus, although we use cross-sectional data, it is possible to presume that victimization experiences, in most cases, precede reported levels of fear or risk.

Variables measuring routine-activity risk factors are also included in the analyses. Exposure, indicating how accessible a target is in relation to motivated offenders, is indicated by three different measures. First, the number of nights during the previous week that the home was left unoccupied (*home unoccupied*) reflects this accessibility, especially in terms of household victimization. The number of attractive *access routes* leading into and out of the respondent's home (e.g., presence of ground floor windows, alley behind the home) is also indicative of exposure, as is whether or not the respondent's home is a *corner residence*, as opposed to situated in the middle of a block.[3]

Another routine activity characteristic, target attractiveness, is also measured in several different ways. Target attractiveness — referring to the value the target represents — is measured first by an index tapping the number of expensive, portable household goods within the respondent's home (*expensive goods*), including a portable color television, a VCR, a 35 mm camera, a home computer, and a bicycle or motorcycle. Secondly, *family income* is used as a measure of target attractiveness.[4]

We also include measures of guardianship, or the ability of targets to effectively prevent victimization through the use of other people or objects. First, we use an index representing the number of *safety precautions* employed by the respondent two years prior to the survey, including use of locks on doors, installation of extra locks (e.g., deadbolt locks or chains), use of a light-timer device or leaving lights on, membership in a crime-prevention program, ownership of a burglar alarm or security system, ownership of a dog, ownership of a weapon, or having neighbors watch property when away. Further, we include a measure of the number of *guardianship barriers* associated with the respondent's home, including the presence of a tall fence or hedge and the presence of an empty house or vacant property next door. The more barriers around a respondent's property, the more likely it is that neighbors will be

TABLE 1: Variables, Metrics, and Descriptive Statistics[a]

Variables	Metrics	Descriptive Statistics Mean	Std. Dev.	Range
Dependent variables				
Burglary-specific fear	(O = no, 1 = yes)	.385	.487	0-1
Perceived risk	(O = safe, 1 = unsafe)	.241	.428	0-1
Explanatory variables				
Individual-level characteristics				
Age	(1 = 10-19 . . . 7 = 70 and older)	4.367	1.717	1-7
Gender	(0 = female, 1 = male)	.503	.500	0-1
Race	(0 = white, 1 = nonwhite)	.150	.358	0-1
Home unoccupied	(nights per week)	1.791	1.969	0-7
Expensive goods	(no. of items owned)	2.533	1.408	0-5
Safety precautions	(no. of precautions)	3.837	1.509	0-9
Live alone	(O = no, 1 = yes)	.258	.437	0-1
Family income	(1 = LT $10k . . . 7 = GT $100k)	3.371	1.370	1-7
Guardianship barriers	(no. of barriers)	.487	.638	0-2
Access routes	(no. of routes)	.333	.543	0-2
Corner residence	(0 = no, 1 = yes)	.423	.495	0-1
Burglary victimization	(0 = no, 1 = yes)	.170	.375	0-1
Contextual characteristics[b]				
Busy places	(no. of places nearby)	3.441	1.340	.82-7.19
Ethnic heterogeneity	(proportion nonwhite * proportion white)	.090	.079	0-.25
Neighborhood incivilities	(no. of indicators)	1.328	.704	0-3.30
Social integration	(no. of indicators)	3.516	.801	1.41-5.35
Burglary rate	([no. reported to police in 1989 and 1990/ households] * 100)	7.463	3.903	1.81-18.52

[a] The total sample size is 5,090 individuals, 300 neighborhoods, and 100 census tracts

[b] Summary statistics for busy places, ethnic heterogeneity, neighborhood incivilities, and social integration are based on the neighborhood block-pair as the unit of analysis; summary statistics for burglary rate are based on the census tract as the unit of analysis.

hindered from providing effective guardianship, thereby indicating low guardianship. A social dimension of guardianship is measured by whether or not the respondent *lives alone.* Sociodemographic variables for *age, race, and gender* are also included as individual-level explanatory variables. We are interested in any differences regarding the explanatory power associated with these individual-level characteristics in understanding risk perception versus burglary-specific fear.

CONTEXTUAL VARIABLES

Several contextual, neighborhood-level attributes are created by aggregating individual responses to certain items pertaining to neighborhood disorder or disorganization within city-block groups. Neighborhood population density or traffic is measured by averaging the number of places for public activity (*busy places*) within the neighborhood block pair as reported by respondents. Busy places include schools, convenience stores, bars, fast-food restaurants, office buildings, parks or playgrounds, shopping malls, hotels/motels, and bus stops. Neighborhood *incivility* is measured by an index averaging the number of problems or signs of disorder (as perceived by individual respondents) within the neighborhood, including the presence of teenagers on the street, litter in the street, abandoned or boarded buildings, poor street lighting, and signs of vandalism (e.g, graffiti). Neighborhood *ethnic heterogeneity* is measured by multiplying the proportion of respondents within each neighborhood block pair who are white and the proportion nonwhite. Neighborhood *social integration* is measured by an index averaging the indicators of cohesion among fellow-residents, including whether or not respondents recognize strangers on their block, have friends or relatives on their block, watch neighbors' property when they are away, borrow items from neighbors, have lunch or dinner with neighbors, and help neighbors with problems.[5]

A final contextual characteristic was created at the census-tract level with Seattle police data. Specifically, a *two-year, tract-specific residential burglary rate* (per 100 households) was calculated for each of the 100 census tracts in the sample by combining official burglary counts for 1989 and 1990 (from police data), dividing by the total number of households (from 1990 census data), and multiplying by 100. As with the individual-level explanatory variables, we seek to discern whether or not the direct and moderating effects of these contextual characteristics are similar in cases of risk perception and burglary-specific fear.

Results for Burglary-Specific Fear

INDIVIDUAL-LEVEL ANALYSES

Given the data and measures described above, an individual-level model as part of an initial random-coefficient regression model for burglary-specific fear (whether or not respondents worry about their homes being broken into at least once a week or more) for individual *i* living in neighborhood *j* and census tract *k* was specified as:

$$\log(p_{ijk}[1-p_{ijk}]^{-1}) = \beta_{0jk} + \beta_{1jk}(age_{ijk}) + \beta_{2jk}(gender_{ijk}) + \beta_{3jk}(race_{ijk}) + \beta_{4jk}(home\ unocc_{ijk}) + \beta_{5jk}(income_{ijk}) + \beta_{6jk}(goods_{ijk}) + \beta_{7jk}(saftey_{ijk}) + \beta_{8jk}(live\ alone_{ijk}) + \beta_{9jk}(barriers_{ijk}) + \beta_{10jk}(access_{ijk}) + \beta_{11jk}(corner_{ijk}) + \beta_{12jk}(victim_{ij}) + e_{ijk} \quad (1)$$

with each of the predictors being grand-mean centered. After initially assuming all coefficients in the above model to be variable across Seattle neighborhoods and census tracts, a simplified model was estimated in which all coefficients that *did not* vary were specified as fixed. The resulting level 2, or between-neighborhood model for the initial random-coefficient regression model for burglary-specific fear of crime is, then, the following:

$$\begin{aligned} B_{0jk} &= \theta_{00k} + u_{0jk} \\ B_{12jk} &= \theta_{120k} + u_{12jk} \\ B_{qjk} &= \theta_{q0k}\ for\ q = 1-11 \end{aligned} \quad (2)$$

and the resulting level 3, or between-tract model is as follows:

$$\theta_{q0k} = \pi_{q00}\ for\ q = 0-12 \quad (3)$$

No coefficients exhibit significant variability across census tracts, as indicated by the between-tract model specified in equation 3, and as the results from the random-effects panel of Table 2 indicate. At the neighborhood level, by comparison, adjusted mean burglary-specific fear varies across neighborhoods, and the effect of burglary victimization on this fear also varies across Seattle neighborhoods.

The fixed-effects panel of Table 2 indicates that nearly all the individual-level predictors included in the model have significant effects on burglary-specific fear. Two of the sociodemographic indicators — age and race — have effects that reach standard levels of significance, although the relationships are opposite in direction to what has been found previously in the fear-of-crime literature. For instance, age is negatively related to fear. The more easily interpreted exponentiated coefficient for age suggests that for each unit increase in age, the odds of experiencing burglary-specific fear decreases by nearly 3%. Further, nonwhites are shown to be less fearful of burglary — nonwhites exhibit approximately 5% reduction in the odds of experiencing fear in comparison to whites. The other sociodemographic variable — gender — is insignificant. These findings seem to completely contradict long-held notions that the elderly, women, and nonwhites are more fearful of crime. Implications of these findings are further discussed below.

Most of the routine activity variables show relationships with burglary-specific fear which indicate that those survey respondents with "more dangerous" lifestyles are more fearful. For example, exposure is positively related to fear: for each unit increase in the access routes and corner residence variables, the odds of experiencing fear increase 4% and 6% respectively. Further, target attractiveness is positively related to fear — the more expensive household goods a respondent owns, the more fearful he/she is of burglary (increasing the odds of fear by nearly 3% for each unit increase in expensive goods). However, the other indicator of target attractiveness — family income — is insignificant in this model. Guardianship seems to have mixed effects on fear. On the one

TABLE 2: Random Coefficient Regression Model for Logistic Regression Coefficients of Burglary-Specific Fear of Crime in Seattle

Fixed effect	Coef.	Std. Error	Exp. (Coef.)	t ratio[b]
Mean fear, π_{000}	.383	.008	1.467	50.144
Age, π_{100}	-.026	.004	.974	-6.001
Gender, π_{200}	.018	.013	1.018	1.340
Race, π_{300}	-.050	.019	.951	-2.592
Home unoccupied, π_{400}	.005	.004	1.005	1.311
Family income, π_{500}	.002	.006	1.002	.395
Expensive goods, π_{600}	.026	.005	1.026	4.693
Safety precautions, π_{700}	.028	.005	1.028	6.018
Live alone, π_{800}	-.011	.017	.989	-.655
Guardianship barriers, π_{900}	.028	.011	1.028	2.671
Access routes, π_{1000}	.039	.013	1.040	3.104
Comer residence, π_{1100}	.059	.014	1.061	4.203
Burglary victimization, π_{1200}	.144	.020	1.155	7.310

Random Effect	Variance Component	Z score	P value
Level 2			
Mean fear, u_{0jk}	.003	2.169	<.05
Burglary victimization, u_{12jk}	.024	2.597	<.05
Level 1			
Level-1 extra binomial error, e_{ijk}	.218	47.779	<.05
Model χ^2	272.74		
Degrees of freedom	11		
Probability	<.05		
(N = 5,090)			

[a] The level 2 unit is the neighborhood.
[b] The t-ratio calculations are based upon the complete estimates for coefficients and standard errors provided by the ML3E program rather than being based upon the rounded figures presented in columns 1 and 2.

hand, the more guardianship barriers a respondent's home has, limiting the potential for effective guardianship, the more fearful he/she is of burglary. On the other hand, safety precautions (increasing guardianship) are positively related to fear. So, increased precautionary measures *and* apparent barriers to effective guardianship lead to *elevated* levels of fear even though safety precautions are negatively related to the risk of actual burglary victimization and guardianship barriers are positively related to such a risk (Rountree 1994).

Finally, *previous* experience with burglary victimization increases fear of household break-ins — the odds increase by about 16% compared to those not experiencing such victimization. According to the significant chi-square statistic, this model with individual-level predictors is a significant improvement over a baseline burglary-specific fear of crime model containing only an intercept (varying across neighborhoods and census tracts) and an individual-level extra binomial error term.[6]

TRACT-LEVEL CONTEXTUAL EFFECTS

We estimated next a contextual model in which neighborhood- and tract-level indicators of crime and disorganization are added to the model in an attempt to account for the variability in adjusted mean burglary-specific fear across neighborhoods and the variability in the effect of burglary victimization on fear across different neighborhoods. The level 2, or between-neighborhood model is thus extended as follows in order to incorporate the contextual characteristics of interest while the between-tract model remains unchanged (since there was no variability at this third level):

Level 2 (between-neighborhood) model (4)

$$B_{0jk} = \theta_{00k} + \theta_{01k} busy\,pl_{jk} + \theta_{02k} incivil_{jk} + \theta_{03k} ethnic_{jk} + \theta_{04k}\, integ_{jk} + u_{0jk}$$
$$B_{12jk} = \theta_{120k} + \theta_{121k} busy\,pl_{jk} + \theta_{122k} incivil_{jk} + \theta_{123k} ethnic_{jk} + \theta_{124k}\, integ_{jk} + u_{12jk}$$
$$B_{qjk} = \theta_{q0k}\, for\ q = 1-11$$

Level 3 (between-tract) model (5)

$$\theta_{q0k} = \pi_{q00}\, for\ q = 0-12$$

At level 2, the variability in burglary-specific fear and the variability in the effect of prior burglary victimization on this fear, then, are estimated as functions of grand means, neighborhood-level busy places, incivilities, ethnic heterogeneity, and social integration. At level 3, all effects are specified as fixed. A single-equation model combining the above revised contextual models with the individual-level model (see equation 1) was estimated. Individual and contextual-level variables that were insignificant were dropped from the model one at a time through the process of backwards elimination, and the results from a reduced contextual model for burglary-specific fear are reported in Table 3.[7]

First, the chi-squared statistic presented in Table 3 indicates that this contextual model is a significant improvement over the previously discussed

TABLE 3: Estimated Effects of Community Characteristics on Logistic Regression Coefficients for Burglary-Specific Fear of Crime in Seattle[a]

Fixed effect	Coef.	Std. Error	Exp. (Coef.)	t ratio[b]
Mean fear				
Base, π_{000}	.388	.007	1.474	53.270
Neighborhood incivilities, π_{020}	.024	.012	1.025	2.090
Neighborhood integration, π_{040}	.022	.010	1.022	2.087
Age, π_{100}	-.029	.004	.971	-6.816
Race, π_{300}	-.059	.019	.942	-3.071
Expensive goods, π_{600}	.027	.005	1.028	5.335
Safety precautions, π_{700}	.027	.005	1.028	5.716
Guardianship barriers, π_{900}	.029	.010	1.030	2.779
Access routes, π_{1000}	.039	.013	1.040	3.082
Corner residence, π_{1100}	.057	.014	1.058	3.994
Burglary victimization				
Base, π_{1200}	.149	.020	1.161	7.307
Neighborhood incivilities, π_{1220}	-.063	.028	.939	-2.246

Random effect	Variance Component	Z score	P value
Level 2			
Mean fear, u_{0jk}	.003	2.018	<.05
Burglary victimization, u_{12jk}	.023	2.530	<.05
Level 1			
Level 1 extra binomial error, e_{ijk}	.218	47.779	<.05
Model χ^2	6.55		
Degrees of freedom	0		
Probability	<.05		
(N = 5,090)			

[a] The level-2 unit is the neighborhood. The t-ratio calculations are based upon the complete estimates for coefficients and standard errors provided by the ML3E program rather than being based upon the rounded figures presented in columns 1 and 2.

model containing only individual-level predictors. The random-effects portion of this table reveals a still-significant variance component for adjusted mean burglary-specific fear at level 2, implying that this fear is still significantly variable across neighborhoods. Apparently, the contextual characteristics that were added could not adequately account for the variability in concern about burglary across neighborhoods. Further, the still-significant variance component for the burglary victimization variable indicates that the added contextual variables were not completely adequate in explaining the variability in the effect of burglary victimization on burglary-specific fear across neighborhoods.

The fixed-effects panel of Table 3 reveals some significant contextual main and moderating effects. Specifically, neighborhood-level incivilities exhibit a significant positive effect on burglary-specific fear. Thus, the positive effect of disorder on our measure of fear is consistent with much previous research despite the fact that most of the extant work uses a non-crime-specific, cognitive, assessment-of-risk measure for fear. In contrast, our finding regarding the main positive effect of neighborhood social integration differs from previous studies using such a measure (e.g., Ferraro 1995; Skogan & Maxfield 1981; Lewis & Salem 1986; Taylor et al. 1984).

Beyond the main effects of context, neighborhood incivilities are also involved in a significant negative interaction with burglary victimization. This interaction indicates that the positive effect of previous burglary victimization on burglary-specific fear is most apparent in relatively orderly neighborhoods (those with fewer incivilities). Highly disordered neighborhoods, on the other hand, temper the effect of burglary victimization, indicating that social context — particularly as it relates to levels of incivility — may have a tendency to "overwhelm" individual-level variables in the explanation of fear.

With respect to the individual-level predictors, the fixed-effects panel of Table 3 shows no new findings; the individual-level relationships are very similar to those found in the initial random-coefficient model for burglary-specific fear (see discussion of Table 2).

Results for Perception of Risk

INDIVIDUAL-LEVEL ANALYSES

The individual-level model for the hierarchical analysis of perception of crime risk is identical to that of the burglary-specific fear of crime (see equation 1), except that p_{ijk} is the probability of perceiving one's neighborhood to be either somewhat or very unsafe from crime. As in the previous random-coefficient regression model, all the coefficients from the individual-level model were initially assumed to vary across Seattle neighborhoods and census tracts, and the coefficients for which this assumption was incorrect were subsequently specified as "fixed." Findings from the resulting initial random-coefficient logistic regression model are summarized in Table 4.[8] The chi-squared statistic for this model suggests that this model is a significant improvement over a baseline risk-perception model.

Significant variance components are reported in the random-effects panel of Table 4 at level 3 for adjusted mean risk perception and at level 2 for adjusted

TABLE 4: Random Coefficient Regression Model for Logistic Regression Coefficients of Crime Risk Perception in Seattle[a]

Fixed effect	Coef.	Std. Error	Exp. (Coef.)	t ratio[b]
Mean risk, π_{000}	.242	.017	1.274	14.526
Age, π_{100}	-.010	.004	.990	-2.726
Gender, π_{200}	-.039	.011	.962	-3.491
Race, π_{300}	-.096	.020	.909	-4.752
Home unoccupied, π_{400}	.004	.003	1.004	1.144
Family income, π_{500}	-.013	.005	.987	-2.732
Expensive goods, π_{600}	.005	.004	1.005	1.097
Safety precautions, π_{700}	.013	.004	1.013	3.259
Live alone, π_{800}	-.003	.014	.997	-.192
Guardianship barriers, π_{900}	.013	.009	1.013	1.460
Access routes, π_{1000}	.040	.011	1.041	3.711
Corner residence, π_{1100}	.017	.014	1.017	1.191
Burglary victimization, π_{1200}	.110	.017	1.116	6.550

Random effect	Variance Component	Z score	P value
Level 3			
Mean risk, r_{00k}	.022	5.620	<.05
Level 2			
Mean risk, u_{0jk}	.007	4.247	<.05
Race, u_{3jk}	.024	3.041	<.05
Home unoccupied, u_{4jk}	.0005	2.029	<.05
Burglary victimization, u_{12jk}	.016	2.532	<.05
Level 1			
Level 1 extra binomial error, e_{ijk}	.141	45.729	<.05
Model χ^2	187.9		
Degrees of freedom	15		
Probability	<.05		
(N = 5,090)			

[a] The level-2 unit is the neighborhood; the level-3 unit is the census tract.
[b] The *t*-ratio calculations are based upon the complete estimates for coefficients and standard errors provided by the ML3E program rather than being based upon the rounded figures presented in columns 1 and 2.

mean risk perception, race, leaving the home unoccupied, and previous burglary victimization. Thus, adjusted mean perception of risk is significantly different across Seattle neighborhoods and census tracts. Furthermore, race, leaving the home unoccupied, and previous burglary victimization have variable effects on risk perception across level 2 units (neighborhoods).

The fixed-effects panel of Table 4 reveals individual-level effects that differ somewhat from those found for burglary-specific fear. Specifically, males are significantly less likely than females to feel danger or unsafe in their neighborhood, but gender was insignificant in the burglary-specific fear model. More consistent with the burglary-specific fear results, findings on race suggest that nonwhites perceive lower risk than their white counterparts and worry less about household break-ins. Age is also negatively related to risk perception, as it was with fear, although the effect appears to be somewhat stronger for burglary-specific fear. In addition to sociodemographic attributes, several other individual-level explanatory variables exert significant main effects on perception of crime and victimization risk. In particular, safety precautions, access routes, and previous burglary victimization are all positively related to risk perception. Thus, the presence of prior victimization experiences and increases in the number of safety precautions and access routes have the effect of heightening perceived risk. Finally, family income is negatively related to perceived risk.

TRACT-LEVEL CONTEXTUAL EFFECTS

The findings for the contextual model for perceived risk are presented in Table 5. The contextual model stems from combining the level 1 model for perceptions of risk with a revised level 2 (between-neighborhood) model that attempts to account for the variability in mean perceived risk and the coefficients for race, home unoccupied, and prior burglary victimization with the addition of neighborhood-level busy places, incivilities, ethnic heterogeneity, and social integration, and a revised level 3 (between-tract) model that attempts to explain the variation in mean perceived risk across tracts with the addition of residential burglary rates. Thus, the revised level 2 and level 3 models are as follows:[9]

Level 2 (between-neighborhood) model (6)

$$B_{0jk} = \theta_{01k} busy\,pl_{jk} + \theta_{02k} incivil + \theta_{03k} ethnic_{jk} + \theta_{04k} integ_{jk} + \theta_{05k} incivil * ethnic_{jk} + u_{0jk}$$
$$B_{3jk} = \theta_{30k} + \theta_{31k} busy\,pl_{jk} + \theta_{32k} incivil_{jk} + \theta_{33k} ethnic_{jk} + \theta_{34k}\, integr_{jk} + u_{3jk}$$
$$B_{4jk} = \theta_{40k} + \theta_{41k} busy\,pl_{jk} + \theta_{42k} incivil_{jk} + \theta_{43k} ethnic_{jk} + \theta_{44k}\, integ_{jk} + u_{4jk}$$

$$\beta_{qjk} = \theta_{q0k}\ for\ q = 1,2,5-11$$

Level 3 (between-tract) model (7)

$$\theta_{00k} = \pi_{000} + \pi_{001} burglary\ rate_k + r_{00k}$$
$$\theta_{q0k} = \pi_{q00}\ \ for\ q = 1-12$$

The purpose of the contextual model resulting from the combining of the level 2 and level 3 models above with the individual-level model for perceived risk is to try to expose the direct and moderating effects of community crime and disorganization on perceived risk. Specifically, we are interested in whether or not the neighborhood-level contextual variables can account for the variability across neighborhoods in adjusted mean risk perception and the relationships between perceived risk and race, home unoccupied, and previous burglary victimization. Likewise, we are interested in whether or not the tract-level burglary rate can account for the variation in adjusted mean risk across Seattle census tracts. The significant chi-squared statistic for this contextual model suggests that it produces an improved fit as compared to the random-coefficient model containing only the individual-level characteristics.

The random-effects panel of Table 5 shows that the contextual model is somewhat successful in that adjusted mean perceived risk is no longer significantly variable across Seattle census tracts. On the other hand, adjusted mean risk perception remains significantly variable across neighborhoods, and the effects of race and burglary victimization on this risk perception are still variable across neighborhood units. The effect of leaving the home unoccupied on this risk, however, is no longer variable across neighborhoods. Thus, the addition of the contextual indicators of crime and disorganization was most effective in accounting for the between-tract differences in adjusted mean perceived risk and between-neighborhood differences in the effects of leaving the home unoccupied on this risk, although it should be noted that most of the variance components in Table 5 represent substantial declines in comparison to those reported in Table 4.

The main and interactive effects for the contextual variables are described in the fixed-effects panel of Table 5. For instance, neighborhood incivility has a positive effect on perceived risk. The significant interaction between incivilities and heterogeneity suggests that the positive effect of incivilities on perceived risk is stronger in ethnically heterogeneous areas. Neighborhood-level community integration is negatively related to perceived risk, and the tract-level residential burglary rate has a positive and significant effect on perceived risk.

Contextual characteristics also exhibit some moderating effects. In such cases, indicators of crime or disorder *condition* the effects of the individual-level attributes. For instance, neighborhood incivilities "contextualize" the effects of race on perceived risk. The significant negative interaction between race and neighborhood incivilities suggests that the finding that nonwhites feel less risk than whites is especially strong in relatively disordered neighborhoods. In very orderly neighborhoods, on the other hand, the negative effect of being nonwhite on risk perception is much weaker. Neighborhood incivilities also condition the effect of leaving the home unoccupied. Exposure by leaving one's home unoccupied at night leads to greater fear, particularly in relatively disorderly neighborhoods. The effect of leaving the home unoccupied is tempered somewhat by relative order and civility within the social context.

The individual-level effects on perceived risk in the contextual model are similar to those found in the estimation of the initial random-coefficient regression model for risk perception (see discussion above). In short, men, nonwhites, and older respondents feel less danger (although age is only

TABLE 5: Estimated Effects of Community Characteristics on Logistic Regression Coefficients for Crime/Victimization Risk Perception in Seattle[a]

Fixed effect	Coef.	Std. Error	Exp. (Coef.)	t ratio[b]
Mean risk				
Base, π_{000}	.227	.008	1.255	27.673
Neighborhood incivilities, π_{020}	.145	.013	1.156	11.187
Neighborhood integration, π_{040}	-.034	.010	.966	-3.521
Incivilities * ethnic heterogeneity, π_{050}	.685	.132	1.984	5.190
Tract-level burglary rate, π_{001}	.014	.002	1.014	6.277
Age, π_{100}	-.006	.003	.994	-1.668
Gender, π_{200}	-.039	.011	.961	-3.623
Race				
Base, π_{300}	-.095	.020	.909	-4.669
Neighborhood incivilities, π_{320}	-.089	.029	.915	-3.124
Home unoccupied				
Base, π_{400}	.003	.003	1.003	.949
Neighborhood incivilities, π_{420}	.013	.004	1.013	3.062
Safety precautions, π_{700}	.014	.004	1.014	3.603
Access routes, π_{1000}	.038	.011	1.039	3.589
Burglary victimization				
Base, π_{1200}	.102	.017	1.107	6.060

Random effect	Variance Component	Z score	p value
Level 3			
Mean risk, r_{00k}	.001	1.490	>.05
Level 2			
Mean risk, u_{0jk}	.004	2.768	<.05
Race, u_{3jk}	.022	2.981	<.05
Home unnocupied, u_{4jk}	.0003	1.662	>.05
Burglary victimization, u_{12jk}	.017	2.648	<.05
Level 1			
Level 1 extra binomial error, e_{ijk}	.141	45.890	<.05
Model χ^2	7.06		
Degrees of freedom	1		
Probability	<.05		
(N = 5,090)			

[a] The level 2 unit is the neighborhood; the level-3 unit is the census tract.
[b] The *t*-ratio calculations are based upon the complete estimates for coefficients and standard errors provided by the ML3E program rather than being based upon the rounded figures presented in columns 1 and 2.

significant at the .10 level in the contextual model), while safety precautions, access routes, and burglary victimization increase perceived risk. The main effect of family income — significant in the random coefficient regression model for perceived risk — is insignificant in the contextual model.

Discussion and Conclusion

The findings from the foregoing analyses suggest that, indeed, there are some important distinctions between a cognitive assessment of risk within one's neighborhood — a measure regularly referred to as "fear" in the fear-of-crime literature — and an emotionally based, crime-specific fear — in particular, burglary-specific fear. These distinctions occur in individual-level effects and contextual effects. Further, micro-macro interactions are distinct for each of the respective concepts.

As alluded to previously, some of the results discussed here contradict earlier work on fear of crime, perhaps because of differences in the concept of *fear*. For instance, our research indicates that younger people and whites are more fearful of burglary, while gender has an insignificant effect on burglary-specific fear. Most of the literature, however, stresses that the elderly, women, and nonwhites fear crime more than other sociodemographic groups. However, most of these previous studies focus on cognitive assessments of neighborhood safety, similar to the risk perception measure used here. Thus, risk perception and actual fear may have different sociodemographic predictors. Partially in support of such a proposition, we find that age is a much more marginally significant predictor of perceived risk, and men are significantly less likely to feel unsafe than women. More consistent with the burglary-specific fear analysis, nonwhites feel safer (less risk) than whites, although in the case of risk perception the effect of being nonwhite interacts with neighborhood incivilities in such a way that nonwhites are particularly less likely than whites to feel unsafe in highly disordered communities. Income — a sociodemographic variable that can also be viewed as a proxy for target attractiveness — appears to be a somewhat better predictor of risk perception in comparison to burglary-specific fear. Though the effect of income on risk perception falls just below the .10 level of significance in the contextual model, it had a significantly negative effect on risk perception in the random coefficient regression model (Table 4). In contrast, income is not a significant predictor of burglary-specific fear, and the direction of the effect is positive. Overall, then, we cannot make the claim that the sociodemographic predictors for fear of crime (household burglary, specifically) and perceived risk are identical.

Beyond sociodemographic findings, other significant microlevel predictors of fear of crime and perceived risk include various routine activities. For the most part, those respondents whose activities left them more vulnerable (i.e., had more expensive goods, more guardianship barriers, access routes, etc.) tended to be more likely to fear crime or feel that their neighborhood was unsafe, although this pattern holds more so for burglary-specific fear than it does for risk perception. Such findings should be stronger for burglary-specific fear, however, given that many of the routine activity measures tap household

vulnerability specifically rather than the vulnerability of a resident in the neighborhood irrespective of residence.

Previous experience with household burglary behaves similarly with respect to both fear and perceived risk — it is a key individual-level predictor of both measures. Previous burglary victimization heightens concern about a repeat burglary (thus increasing burglary-specific fear) as well as increasing the likelihood that one will evaluate one's neighborhood as unsafe. The positive effect of burglary victimization on burglary-specific fear is particularly strong in relatively orderly neighborhoods.

The individual-level predictors discussed above also exhibit differing degrees of variability across contexts in their effects on burglary-specific fear versus perceived risk. Our results showed, for instance, that adjusted mean risk perception varied across tracts and neighborhoods, while adjusted mean burglary-specific fear varied only across neighborhoods. In short, the variation in burglary-specific fear seems to be partitioned quite differently across contextual units than is the variation in perceived risk. The relationships between risk perception and the race and home unoccupied variables were variable across neighborhoods, whereas neither of these variables had differential effects on burglary-specific fear across neighborhoods. Previous burglary victimization, however, did vary in its effects on both fear and perceived risk. Notice that contextual variables interact uniquely with individual-level predictors for burglary-specific fear and risk perception. This interaction is related to the differences in the variability across contexts in individual-level predictors.

We found both similarities and differences among risk perception and burglary-specific fear concerning the main effects of contextual factors. First, neighborhood incivilities are positively related to both burglary-specific fear and risk perception, although this effect appears much stronger for fear than for perceived risk, affirming, to a certain extent, previous work by LaGrange et al. (1992).

An interesting difference between the contextual predictors of burglary-specific fear and perceived risk occurs in relation to the social integration measure. Our findings show that neighborhood cohesion is significantly and negatively related to perceived risk, indicating that social integration serves to diminish residents' perceptions of the neighborhood being unsafe. The relationship between integration and burglary-specific fear, however, is significant and positive. In brief, social integration is found to influence positively fear about household burglary, yet it is negatively related to perceived risk. It is plausible that residents of a relatively close-knit community would feel better about the safety of their neighborhood as a whole — they are probably more likely to feel as if there is adequate informal social control and social support, and this perception is perhaps reflected in a more positive evaluation of the neighborhood's safety. However, community cohesiveness is positively associated with neighborhood socioeconomic status (SES).[11] Perhaps due to the often-found positive relationship between high income and burglary victimization risk, then, residents of integrated communities are more fearful of burglary specifically, though they perceive their neighborhood, overall, to be relatively safe from crime or disorder in general — if these people are going to become victimized,

it is more likely to be of the property-victimization variety. On the other hand, residents of less integrated areas that have more "pressing" victimization concerns; burglary is perhaps the least of their worries.

The literature suggests that residents of integrated communities are more informed about neighborhood events and issues because news and information are well disseminated among cohesive neighbors (Liska & Baccaglini 1990). If the news shared includes information on crime in the neighborhood — occurrences that are more likely to be nonviolent in integrated, prosperous areas — burglary-specific fear levels may rise precisely *because* the community is well integrated. Thus, integration may be positively related to burglary-specific fear because those communities likely to fear burglary — higher-income areas — are most well equipped structurally to disseminate information on break-ins, thereby increasing residents' concerns.

On the other hand, lack of social integration creates an increased sense of danger among residents in the community. Poor integration, along with other indicators of disorganization often plaguing communities in which people feel unsafe (i.e., high crime rates, many incivilities, etc.), serves to increase perceptions of danger by delineating the disorganized state of the neighborhood. In poorly integrated neighborhoods, residents are lacking a strong potential source of social support and informal social control and are thus more likely to negatively evaluate their community's overall safety.

Finally, tract-level burglary rates have a positive impact upon perceived risk. The effect of burglary rates on burglary-specific fear was not estimated due to the insignificant variability in adjusted mean burglary-specific fear across census tracts. Despite the fact that residents of prosperous areas are more likely to experience burglary than, say, violent crime, burglary is also high in poor areas. High burglary rates in either type of neighborhood appear to increase the perception of risk. Consequently, high burglary rates in high-SES neighborhoods could perhaps serve to *buffer* the risk-reducing effects of low incivilities and high integration, whereas high burglary rates in low-SES neighborhoods *reinforce* the effects of other indicators of disorganization likely found in such areas.

In sum, though there were some similarities between risk perception and fear of crime, our findings generally provide further empirical support for the notion that fear and perceived risk represent conceptually distinct constructs. Most previous research ignores substantively important distinctions between different reactions to crime, mainly due to the nearly exclusive focus on measures that tap judgments about risk or assessments of the potential for experiencing crime. However, consistent with recent work in the fear-of-crime tradition that has begun to examine risk as only one of several components of fear (e.g., see Warr & Stafford 1983; Warr 1987; LaGrange, Ferraro & Supancic 1992 — all reviewed earlier), we find that the predictors of traditional measures of fear (risk) can be very different from predictors of an emotional response to burglary victimization. Future work investigating issues surrounding fear of crime should, therefore, continue to be sensitive to the measures used, bearing in mind that a very different set of factors may explain perceived risk or feelings of "urban unease" as opposed to actual crime-specific fear.

Notes

1. We did not delete all cases with missing data on income. Instead, we imputed missing values on income using a group-mean substitution method. Those observations with responses missing for income were assigned a value equal to the mean score for their particular city-block pair.

2. Both the risk perception and fear of crime measures have four-item ordinal response scales associated with their respective items in the Seattle survey. Although multilevel modeling procedures recently have been extended to ordinal logistic regression, in this paper we wanted to present multilevel models most familiar to users and readers. Thus, we decided to dichotomize the response categories, so that the resulting binary response variable could then be estimated with a standard hierarchical logistic model. Ordinal logit models (nonhierarchical) were estimated in order to discern the most appropriate way in which to split the response categories into a dichotomy (Winship & Mare 1984). Results indicated that burglary-specific fear was best dichotomized into groups of worrying about a break-in once a week or daily compared to worrying about being burgled once a month or less. Further, the results indicated that the measure for risk perception was best split into categories of perceiving the neighborhood to be either somewhat or very unsafe as opposed to somewhat or very safe.

3. See Brantingham and Brantingham (1981) or Felson (1994) for a more thorough discussion of the relationship between crime and exposure as related to the ecology of crime.

4. An anonymous reviewer of an earlier draft of this article expressed concern over the possible redundancy of our measures for target attractiveness. The correlation between family income and expensive goods is .409. We estimated models for both fear and perceived risk with (1) both expensive goods and family income in the models, (2) with expensive goods but without family income, and (3) with family income but without expensive goods. The coefficients were very similar across all models, and the significance of none of the variables changed. Since it appeared as if results were not being affected by multicollinearity, we have chosen to include both target attractiveness measures in the models presented here.

5. Since some scholars are critical of using aggregated individual-level data as contextual data, we have compared logistic regression models with aggregated individual-level data (at the tract level) to models with comparable census measures (also at the tract level) and have found that we get very similar results. The only discrepancies occur between models with different contextual variables tapping population age structure and population mobility (e.g., the aggregated measure for tract-level percent aged 17-29 was significant while the census measure for percent aged 17-29, in another model, was insignificant and the aggregated measure for residential mobility was significant while the census measure, in another model, was not). These two discrepancies make more sense if one considers the fact that the sample under-represents young adults and movers. Given the overall consistency in our comparisons, however, and the fact that the sample was not biased in a way which could conceivably affect our aggregated individual-level data (we don't use mobility or age-structure contextual measures), we feel confident that the aggregated survey measures serve as reasonable proxies for truly contextual data.

6. The purpose of this individual-level extra binomial error term is to allow for the possibility that the variances of the individual-level sample fear response probabilities are not distributed precisely as a binomial variable on the basis of the expected values of the estimated logistic regression model. In general, if the variance component corresponding to this term is estimated to be greater (lesser) than 1.0, then the variance in the fear responses at the respondent level is greater (lesser) than would be expected if the conditional probability of the fear responses followed the pure binomial distribution of the standard logistic regression model. The estimates in Table 2 indicate that this error-term variance is statistically significant and thus that there is evidence for extra-binomial variance at the individual level.

7. We recognize that it is standard in hierarchical modeling procedures to attempt to separate the distinct effects of individual-level sociodemographic variables from compositional effects by including compositional measures (e.g., percent male, percent nonwhite, mean age, mean income) in the contextual model. In models not shown here, we have followed such a strategy. However, significant problems with multicollinearity arise when including such compositional

measures at both the neighborhood and tract levels as well as disorganization-related contextual variables. Further, we found the incorporation of the compositional variables to have little effect on the other effect coefficients and thus chose to present the simpler contextual model, without compositional measures, in this paper. However, since we know that the coefficients are virtually unaffected by the inclusion of compositional measures, we feel confident that individual-level effects discussed in our models represent real individual-level effects rather than unmeasured compositional differences between neighborhoods or census tracts.

8. As in the case of the burglary-specific fear models, all variables in the analyses of perception of risk are grand-mean centered.

9. We also considered the possibility that there may be interactions between several of our contextual characteristics. While such interactions were generally insignificant, note that we did include an interaction term for incivilities * heterogeneity in order to explain variability in the intercept (adjusted mean perceived risk, in this case).

10. Remaining inconsistencies between our sociodemographic findings and previous "fear" research (after taking into account differences in measures) could be partially attributed to the fact that our findings are net of the effects of victimization risk factors, such as exposure, target attractiveness, etc. Many of the early studies did not account for such factors. See Stafford and Galle (1984) for a discussion of potential pitfalls in failing to account for exposure to risk.

11. Neighborhood-level social integration and neighborhood-level income have a zero-order correlation of .583.

References

Baumer, Terry L. 1979. "Research on Fear of Crime in the United States." *Victimology* 3:254-63.

———. 1985. "Testing a General Model of Fear of Crime: Data from a National Sample." *Journal of Research in Crime and Delinquency* 22:239-55.

Brantingham, Paul J., and Patricia L. Brantingham (eds.). 1981. *Environmental Criminology*. Sage Publications.

Bryk, Anthony S., and Stephen W. Raudenbush. 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications.

Bursik, Robert J., Jr., and Harold G. Grasmick. 1993. *Neighborhoods and Crime: The Dimensions of Effective Community Control*. Lexington Books.

Clarke, Alan H., and Margaret J. Lewis. 1982. "Fear of Crime among the Elderly." *British Journal of Criminology* 22:49-62.

Felson, Marcus. 1994. *Crime and Everyday Life*. Thousand Oaks Press.

Ferraro, Kenneth F. 1995. *Fear of Crime: Interpreting Victimization Risk*. SUNY Press.

Ferraro, Kenneth, and Randy L. LaGrange. 1987. "The Measurement of Fear of Crime." *Sociological Inquiry* 57:70-101.

Garofalo, James. 1979. "Victimization and the Fear of Crime." *Journal of Research in Crime and Delinquency* 16:80-97.

Garofalo, James, and John Laub. 1979. "The Fear of Crime: Broadening our Perspective." *Victimology* 3:242-53.

Goldstein, Harvey. 1987. *Multilevel Models in Educational and Social Research*. Oxford University Press.

Hox, Joop J., and Itag G. Kreft. 1994. "Multilevel Analysis Methods." *Sociological Methods and Research* 22:283-99.

LaGrange, Randy L., Kenneth Ferraro, and Michael Supancic. 1992. "Perceived Risk and Fear of Crime." *Journal of Research in Crime and Delinquency* 29:311-34.

Lewis, Dan A., and Michael G. Maxfield. 1980. "Fear in the Neighborhoods: An Investigation of the Impact of Crime." *Journal of Research in Crime and Delinquency* 17:160-89.

Lewis, Dan A., and Greta Salem. 1986. *Fear of Crime: Incivility and the Production of a Social Problem*. Transaction Books.

Lindley, Dennis V., and A.F.M. Smith. 1972. "Bayes Estimates for the Linear Model (with discussion)." *Journal of the Royal Statistical Society, Series B* 34:1-41.

Liska, Allen E., and William Baccaglini. 1990. "Feeling Safe by Comparison: Crime in the Newspapers." *Social Problems* 37:306-74.

Liska, Allen E., Joseph J. Lawrence, and Andrew Sanchirico. 1982. "Fear of Crime as a Social Fact." *Social Forces* 60:760-70.

Liska, Allen E., and Barbara D. Warner. 1991. "Functions of Crime: A Paradoxical Process." *American Journal of Sociology* 96:1441-63.

Mason, William W., George Y. Wong, and Barbara Entwisle. 1983. "Contextual Analysis through the Multilevel Linear Model." *Sociological Methodology* 1983-84:72-103.

Miethe, Terance D., and David McDowall. 1993. "Contextual Effects in Models of Criminal Victimization." *Social Forces* 71:741-59.

Miethe, Terance D., and Robert F. Meier. 1994. *Crime and Its Social Context: Toward an Integrated Theory of Offenders, Victims, and Situations*. SUNY Press.

Ortega, Suzanne T., and Jessie L. Myles. 1987. "Race and Gender Effects on Fear of Crime: An Interactive Model with Age." *Criminology* 25:133-52.

Prosser, Robert, J. Rasbash, and H. Goldstein. 1991a. *ML3: Software for Three-Level Analysis*. London: Institute of Education, University of London.

____. 1991b. *Data Analysis with ML3*. London: Institute of Education: University of London.

Raudenbush, Stephen W. 1984. Applications of a Hierarchical Linear Model in Educational Research. *Dissertation Abstracts International* 45:1713A.

Ringdal, Kristen. 1992. "Recent Developments in Methods for Multilevel Analysis." *Acta Sociologica* 35:235-43.

Rountree, Pamela Wilcox. 1994. *The Antecedents and Effects of Criminal Victimization in Seattle: A Multi-Level Contextual Analysis*. Ph.D. diss., Duke University.

Rountree, Pamela Wilcox, and Kenneth C. Land. 1996. "Burglary Victimization, Perceptions of Crime Risk, and Routine Activities: A Multilevel Analysis across Seattle Neighborhoods and Census Tracts." *Journal of Research in Crime and Delinquency* 33:147-80.

Rountree, Pamela Wilcox, Kenneth C. Land, and Terance D. Miethe. 1994. "Macro-Micro Integration in the Study of Victimization: A Hierarchical Logistic Model Analysis across Seattle Neighborhoods." *Criminology* 32:387-414.

Skogan, Wesley G., and Michael G. Maxfield. 1981. *Coping with Crime: Individual and Neighborhood Reactions*. Sage.

Stafford, Mark C., and Omer R. Galle. 1984. "Victimization Rates, Exposure to Risk, and Fear of Crime." *Criminology* 22:173-85.

Strenio, J.L.F., H.I. Weisberg, and A.S. Bryk. 1983. "Empirical Bayes Estimation of Individual Growth Curve Parameters and Their Relationship to Covariates." *Biometrics* 39:71-86.

Taylor, Ralph B., Stephen D. Gottfredson, and Sidney Brower. 1984. "Block Crime and Fear: Defensible Space, Local Social Ties, and Territorial Functioning." *Journal of Research in Crime and Delinquency* 21:303-31.

Taylor, Ralph B., and Margaret Hale. 1986. "Testing Alternative Models of Fear of Crime." *The Journal of Criminal Law and Criminology* 77:151-89.

van der Wurff, Adri, Leendert van Staalduinen, and Peter Stringer. 1989. "Fear of Crime in Residential Environments: Testing a Social Psychological Model." *The Journal of Social Psychology* 129:141-60.

Warr, Mark. 1987. "Fear of Victimization and Sensitivity to Risk." *Journal of Quantitative Criminology* 3:29-46.

Warr, Mark, and Mark Stafford. 1983. "Fear of Victimization: A Look at the Proximate Causes." *Social Forces* 61:1033-43.

Wilson, James Q., and George L. Kelling. 1982. "Broken Windows." *The Atlantic Monthly* March:29-38.

Winship, Christopher, and Robert D. Mare. 1984. "Regression Models with Ordinal Variables." *American Sociological Review* 49:512-25.

Self-Determination as a Source of Self-Esteem in Adolescence*

TIMOTHY J. OWENS, *Indiana University at Indianapolis*
JEYLAN T. MORTIMER, *University of Minnesota*
MICHAEL D. FINCH, *University of Minnesota*

Abstract

This research examines the effects of the sense of self-determination in three spheres of male adolescents' lives (the family, the school, and the workplace) on positive self-esteem, using five waves of panel data from the Youth in Transition study (grade 10 to five years after high school). Though the self-concept literature has emphasized social sources of self-evaluation, these analyses point to the importance of generalization and attribution processes. It is concluded that the same kinds of experiences that have been found to foster positive psychological outcomes in adulthood — personal discretion in decision making and interesting, challenging tasks — have developmental significance in adolescence as well.

The development of the self-concept has long fascinated social scientists, since Cooley (1902) formulated the "looking glass self" and James (1892) and Mead (1934) distinguished the "I" and the "me." In some respects, the individual's self-image is a direct consequence of interpersonal processes, as suggested by Rosenberg's (1981) principles of "reflected appraisals" and "social comparisons." While social sources of the self-image have received predominant emphasis in the self-concept literature (Gecas & Schwalbe 1983; Wylie 1979), these are not the only bases of self-image formation.

Rosenberg's principle of self-attribution (1981) posits that persons form conclusions about themselves by observing their own performances and attainments. Though the standards for such self-evaluations derive from social values and expectations, the direct mediation by others in producing these

** This research was supported by grants from the National Institute on Aging (AG03325), the University of Minnesota Computer Center, and the IUPUI Office of Integrated Technology. The authors would like to thank Geoffrey Maruyama for his comments and advice throughout the data analysis, Melvin Kohn for several very helpful suggestions on an earlier draft, and the participants in the Indiana University Program Seminar on Identity, Self, Role, and Mental Health. Portions of this article were presented at the annual meetings of the American Sociological Association, Washington, D.C., August 1995.* *Direct correspondence to Timothy J. Owens, Department of Sociology, Indiana University at Indianapolis, 425 University Blvd., Indianapolis, IN 46202-5140. E-mail: TOWENS@INDYVAX.IUPUI.EDU.*

judgments is unnecessary. The meaning assigned to one's observations of self is of foremost importance (Rosenberg 1981; see also Markus, Cross & Wurf 1990; Owens 1993). Achievements will be seen as reflecting the self when they are internally caused or motivated, not the result of external forces. Rosenberg and Pearlin (1978) use the principle of self-attribution to explain the absence of a relationship between social class and self-esteem among children. For children, socioeconomic status is ascribed, reflecting the parent's accomplishments. According to the investigators, the child's own achievement influences the developing self. Consistently, psychological research, grounded in self-perception (Bem 1972) and self-attribution (Kelley & Michela 1980) theories, has shown that when people conceive their behavior as internally motivated, as determined by their own volition, there are stronger consequences for the self (Condry 1977; Deci 1975; Weiner 1979).[1]

Much research on adult self-concept development emphasizes the importance of achievements and attainments (Andrisani & Nestel 1976; Cohn 1978; Elder 1969, 1974; Goodwin 1977; Gurin & Gurin 1976; Jacques & Chason 1977; Luck & Heiss 1972; Owens 1992a). Kohn and Schooler (1973, 1983; Kohn 1969) have elucidated the causal processes involved by identifying the components of social class position that influence psychological functioning. Their analyses show that experiences of self-direction at work are of crucial importance (Kohn & Schooler 1983; Slomczynski, Miller & Kohn 1981). Other investigators have similarly emphasized autonomy as a source of self-esteem and a sense of competence (Franks & Marolla 1976; Gecas & Schwalbe 1983; Korman 1970; Mortimer, Finch & Kumka 1982; Mortimer & Lorence 1979b; Spenner & Otto 1985). Research on the effects of self-direction have been extended to housework (Schooler et al. 1984; see also Olsen 1974) and to the schoolwork of children (Miller, Kohn & Schooler 1982).[2]

According to Kohn and Schooler (1983), the effects of work conditions on the personality can be explained by a generalization process: "a direct translation of the lessons of the job to outside-the-job realities" (290). Kohn's generalization model is similar to Rosenberg's principle of self-attribution, though the latter is formulated more specifically with reference to the self-concept. For both, psychological development occurs through the resolution of important problems (Kohn 1976; Kohn & Schooler 1973; Rosenberg 1981). If individuals experience self-direction, successfully adapt to the challenges facing them, and attribute their successes to themselves, they will come to view themselves as capable and worthy.

In the present study, we extend these investigations to the period of adolescence, using five waves of panel data from Bachman's (Bachman, O'Malley & Johnston 1978) Youth in Transition study. We examine the implications of a sense of self-determination and control for the development of self-esteem in three central spheres of adolescents' lives: the family, the school, and for those who hold part-time jobs while still in school, the workplace. Further, it is worth noting that while many authors have discussed the importance of adolescence in developing a basic sense of competence, efficacy, and worth (e.g., Clausen 1991, 1993; Markus, Cross & Wurf 1990; Owens 1994), linkages between the macro aspects of institutional context and self-esteem have been neglected (see Owens 1992a for an attempt to link social context to self-

esteem). In the present study we aim to bridge this gap by explicitly examining whether the perception of freedom and autonomy in three key adolescent spheres of development — the family, the school, and the workplace — are generalized and ultimately attributed to the self, thereby contributing to a sense of self-esteem or worth. In so doing, we take an underlying ecological perspective (Bronfenbrenner 1979). That is, we seek to understand whether subjective experiences of control and freedom in three macro spheres influence the developing individual.

Our theoretical rationale for linking self-attribution and generalization processes is aided by self-determination theory (Deci, Eghrari & Patrick 1994; Deci & Ryan 1985). According to this view, the nature of involvement in, and perception of, a social context influences the internalization and integration of values and regulatory processes associated with that context. As the "person identifies with the value of an activity and accepts full responsibility for doing it . . . one's behavior emanates from one's self; it is self-determined" (Deci, Eghrari & Patrick 1994:121). Expanding this process to the present study, the greater the *perception* of self-determination in the three contexts under examination, the more positive the impact on a youth's sense of self-esteem and worth.

Turning first to the family context, research on the effects of parental control on adolescent development is quite consistent with the principles of attribution and generalization. It appears that the optimal control strategy, fostering a wide range of positive child outcomes (including self-esteem), is participative (Bowerman & Elder 1964; Elder 1962, 1963, 1968; Kandel & Lesser 1972; Lamborn et al. 1991; Rollins & Thomas 1979). The parent shares decision making with the adolescent. The use of reasoning, the information provided by parents, and the child's active participation in decision making foster a sense of efficacy in coping with problems (Barber & Rollins 1990; Rollins & Thomas 1979; Steinberg, Mounts & Lamborn 1991). Moreover, external coercive control generally has negative consequences (Clausen 1993; McLeod & Shanahan 1993), while democratic control practices "induce within the child internalized motivation for his behavior" (Rollins & Thomas 1979; see also Dornbusch 1989; Lamborn et al. 1991).

Since much of the adolescent's time is spent in school, the structuring of the educational environment could also have implications for the developing self-image (Covington 1992). Self-directed school experiences, like democratic patterns of parental control, may foster internally motivated behavior.[3] When the school environment provides a context for self-initiated activities, there are stronger and potentially more positive implications for the self than under conditions of external control (Weiner 1979; see also Covington 1992).[4] Intrinsic motivation has also been found to heighten academic achievement (Condry 1977; Weiner 1979), which has similar positive consequences for the self (Brookover, Thomas & Paterson 1964; Calsyn & Kenny 1977; Maruyama, Finch & Mortimer 1985; Purkey 1970; Rogers, Smith & Coleman 1978). We therefore expect that if the student develops an intrinsic motivation toward schoolwork, these experiences will have positive implications for self-esteem.

The workplace is increasingly a third institutional context for personal development in adolescence (Greenberger & Steinberg 1986; Mortimer et al. 1992). In the Youth in Transition panel, 35% of the students were employed in

the tenth grade and 55% by the end of the twelfth. Whereas some commentators have been optimistic about the implications of work activities for adolescent development (Coleman 1974; Finch et al. 1991), others are more pessimistic (Steinberg, Fegley & Dornbusch 1993; Steinberg, Mounts & Lamborn 1991). Analyses of the Youth in Transition data suggest that long hours of work may depress academic achievement (Finch & Mortimer 1985; Mortimer & Finch 1986b), but assessment of a more recent cohort, in the Youth Development study, indicates that work intensity has little developmental effect (Mortimer, Shanahan & Ryu 1993). However, it must be remembered that adolescent employment, like that of adults, differs on several dimensions (Bachman, O'Malley & Johnston 1978; Greenberger, Steinberg & Ruggiero 1982); it is increasingly evident that the full impacts of working depend on the nature of that experience (Finch et al. 1991; Mortimer, Shanahan & Ryu 1993; Mortimer & Shanahan 1994; Shanahan et al. 1991). Extending this work, we investigate the effects of intrinsic motivation in work on the development of self-esteem.

In summary, this study assesses the implications of the sense of self-determination in three spheres — the family, the school, and the workplace — for the developing adolescent self-image. In doing so, we extend Rosenberg's self-attribution and Kohn's generalization theories to adolescent development.

Data Source

The data are from Bachman's (1974) Youth in Transition study, a five-wave longitudinal study with a multistage sampling design representing 2,213 tenth-grade boys attending 87 public high schools in the continental U.S. in the fall of 1966. We use data from all five waves: wave 1 (tenth grade, approximately 15 years old), wave 2 (eleventh grade, 17 years old), wave 3, (twelfth grade, 18 years old), wave 4 (one year after high school, 19 years old), and wave 5 (five years after high school, 23 years old). Wave 1 data were collected in the fall in the subjects' schools; subsequent waves were collected in the spring at such "neutral sites" as libraries, community centers, and churches, except for the final wave, which was by mailed questionnaire. (The original sample was weighted to correct for school size. All analyses reported here are based on the weighted panel, N = 2,154 in grade ten.)

Although attrition is a major problem in panel studies, 73.5% of the original study participants were included in the final wave (Bachman, O'Malley & Johnston 1978). There were greater losses among African Americans, dropouts, persons of lower socioeconomic background, and persons of lower intelligence.[5] To assess the severity of these losses, Bachman and his colleagues performed an analysis of the relations among the wave 1 variables for the wave 1 and wave 5 respondents. Judging from the similarity of the correlations, they concluded that biases resulting from sample attrition are not serious, particularly if the intent is to estimate relations among variables, not the absolute size of population parameters. Moreover, analyses of data from a control group, queried only at times 1 and 4, indicate no systematic biases from repeated measurements.

Given the relatively small number of African Americans (N = 98), especially in the fifth wave, our analyses are confined to the white panel members. Further

justifying their exclusion, Bachman, O'Malley, and Johnston (1978) caution that because the African American students were highly concentrated in a small number of schools, what appear to be racial differences may actually be school effects. Furthermore, our analyses exclude persons with missing data on any of the variables in each model. We study the remaining pool of subjects through all five waves of data collection. (For the family model 965 respondents provided the data and 967 for the school model.)[6] In estimating the work model, we restrict the analysis to students who were working during the eleventh and twelfth grades ($N = 222$). This restriction focuses attention on those who manifest a rather strong attachment to the labor force.

Bachman, O'Malley & Johnston (1978) find that adolescent self-esteem is positively related to the quality of family relations, positive school attitudes, and ambitious job attitudes, thus showing that attitudes about these important life domains and the self are intertwined. However, Bachman has not focused directly on self-determination in these spheres of life as sources of self-esteem.

Conceptualization and Measurement of the Constructs

AUTONOMY IN FAMILY, SCHOOL, AND WORKPLACE

Research on social structure and personality has pointed to autonomy, or specifically to occupational self-direction, as a central determinant of adult psychological change (House 1980). Kohn and Schooler's (1982:1259) approach emphasizes the job's objective structure as providing opportunities for "the use of initiative, thought, and independent judgment." Mortimer and her colleagues (Mortimer & Finch 1986a; Mortimer & Lorence 1979a, 1979b; Mortimer, Lorence & Kumka 1986) have used the term *work autonomy* to refer to other experiences of work that are implied by "occupational self-direction," including the subjective appraisal of challenge, with findings that closely parallel Kohn and Schooler's work. But the level of congruence between objective and subjective job assessments has been the subject of some controversy (House 1981; O'Brien 1986). Some argue that the objective conditions of work (optimally measured by a job analyst) are what really matter for psychological change (Kohn 1980, 1981). Others take the position that the person's subjective perception is of primary importance (e.g., Griffin 1983; Pfeffer 1981). Research on job satisfaction (Mirowsky 1987; O'Brien 1986; Seashore & Taber 1975) and health (House 1981) suggests that the perception of one's job attributes is a more potent influence on the response to work than the objective situation. But this may not be the case for other psychological phenomena, such as self-esteem.

We view the person's sense of self-determination as one subjective outcome of what Kohn and Schooler (1983) call self-direction. If the structure of activities in the family, school, or workplace allows independent judgment, one would likely perceive one's acts as self-determined and controlled. Complexly structured activities would generally be more interesting and challenging, stimulating an intrinsic motivation to engage in them. Thus, a sense of self-determination may arise from the amount of autonomy permitted by the environment or by the attractive and engaging features of the activity itself. It is conceivable that even in a highly structured and authoritarian classroom

situation, a student may find the subject matter of such inherent interest that schoolwork is viewed as internally motivated, stemming from one's own desire to learn. Though we believe that substantively complex and autonomous environments foster a sense of self-determination and intrinsic motivation, we hypothesize that the subjective phenomenon itself will have positive consequences for self-esteem, irrespective of its sources.

Surely, the ideal study would have measures both of the objective environment and of subjective perceptions to allow assessment of the relationships between them and their effects on self-esteem. The objective structuring of the situation may influence self-esteem and other psychological variables mainly indirectly, through the subjective phenomena of self-determination and control. Still, a number of researchers contend that in order to understand how social contexts influence development and change, it is imperative to gauge the individual's subjective assessments and perceptions of his or her social environment and relationships with others (see Mortimer & Lorence 1995).

In the present study, only self-report data are available, thus precluding truly objective measurement. Furthermore, the questions asked pertain more to the respondent's feelings of intrinsic motivation and control than to the amount of autonomy inherent in the structure of activities or to the character of the environment. This is especially true of the school and work contexts. The family measures focus on decision-making processes and thus more closely reference the environmental structure. All measures appropriately address the sense of self-determination in important spheres of adolescents' lives.

We performed a confirmatory factor analysis using LISREL 7 (Jöreskog & Sorböm 1989) to estimate parameters and to test the overall goodness of fit of the three context-specific self-determination measurement models (as well as the self-esteem measurement model). Assessment of fit helps determine whether a given theoretical model adequately describes the pattern of relationships within a set of data.[7]

Self-Determination in the Family

Four questions related to participation in decision making were used to measure the sense of self-determination in the family. All questions were measured on a 5-point scale (the first and last response options are listed in parentheses below each item).

(1) How much influence do you feel you have in family decisions that affect you? (A great deal . . . Little or no influence)
(2) How often do your parents listen to your side of the argument? (Always . . . Never)
(3) How often do your parents talk over important decisions with you? (Always . . . Never)
(4) How often do your parents act fair and reasonable in what they ask of you? (Always . . . Never)

Bachman (1970:20) labeled these items "amount of reasoning with son." If adolescents feel they influence decisions that affect them, if parents listen, talk over decisions, and are "reasonable," we might conclude that the adolescents have some control in the family domain. As Figure 1 indicates, the model fits

the data well, with a fit ratio of 3.47 ($\chi^2 = 52.02$, df = 15) and a goodness-of-fit index (GFI) of .993.

Self-Determination in School

While the items for the family domain focus on patterns of decision making, the measures in the school sphere express the perception that one's activities in school are internally directed and controlled. If schoolwork is seen as internally motivated, determined by one's own volition because of its "special meaning" and "interest," or because it gives "a chance to learn" or to improve one's "ability to think and solve problems," it will have positive implications for the self (Condry 1977; Covington 1992; Deci 1975; Rosenberg 1981; Weiner 1979).[8] Again, all items were measured on a 5-point scale, with the first and last response options listed in parentheses.

(1) I put a great deal of myself in some things at school because they have special meaning or interest for me. (Very much . . . Not at all)
(2) I enjoy school because it gives me a chance to learn many interesting things. (Very much . . . Not at all)
(3) I like school because I am improving my ability to think and solve problems. (Very much . . . Not at all)
(4) How interesting are most of your courses to you? (Very exciting and stimulating . . Very dull)

The measurement model for school self-determination (Figure 2) also fits the data reasonably well, with a fit ratio of 4.72 ($\chi^2 = 202.82$, df = 43) and a GFI of .977.

Self-Determination in the Workplace

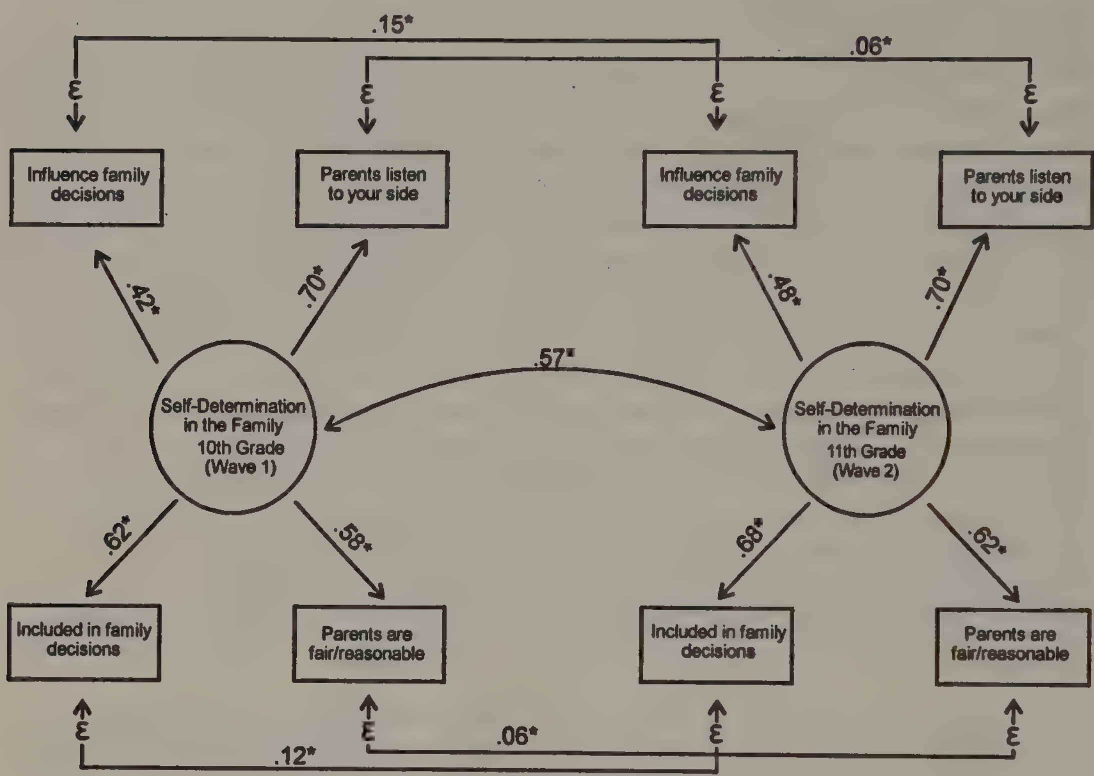
In the work sphere, the first item taps the nature of supervision. The others pertain to one's subjective orientation to the work itself. All items were again measured on a 5-point scale, with the first and last response options listed in parentheses.

(1) How often do you get a chance to work with a supervisor in planning what your work will be — like what you will be doing, or how you should do it? (Almost always . . . Never)
(2) How often are you interested enough to do more work than your job requires? (Most of the time . . . Never)
(3) How often do you feel the work you are assigned on your job is meaningful and important? (Almost always . . . Never)

The more the job is perceived as "interesting," "meaningful," and "important," the more it will provide the basis for a sense of intrinsic motivation and control. These three measures were positively related to a fourth, "How much does your job give you a chance to be independent?" Because this item was missing in the third wave of data collection, however, it was not included in the construct. The measurement model for work self-determination (Figure 3) fits the data well, with a fit ratio of 1.34 ($\chi^2 = 6.71$, df = 5) and a GFI of .994.

Despite the character of jobs typically held by teenagers (almost three-quarters of the employed eleventh-graders in the panel worked as operatives,

FIGURE 1: Measurement Model of Self-Determination in the Family



$\frac{\chi^2}{df} = \frac{52.02}{15} = 3.47$ GFI = .993 *$p<.05$

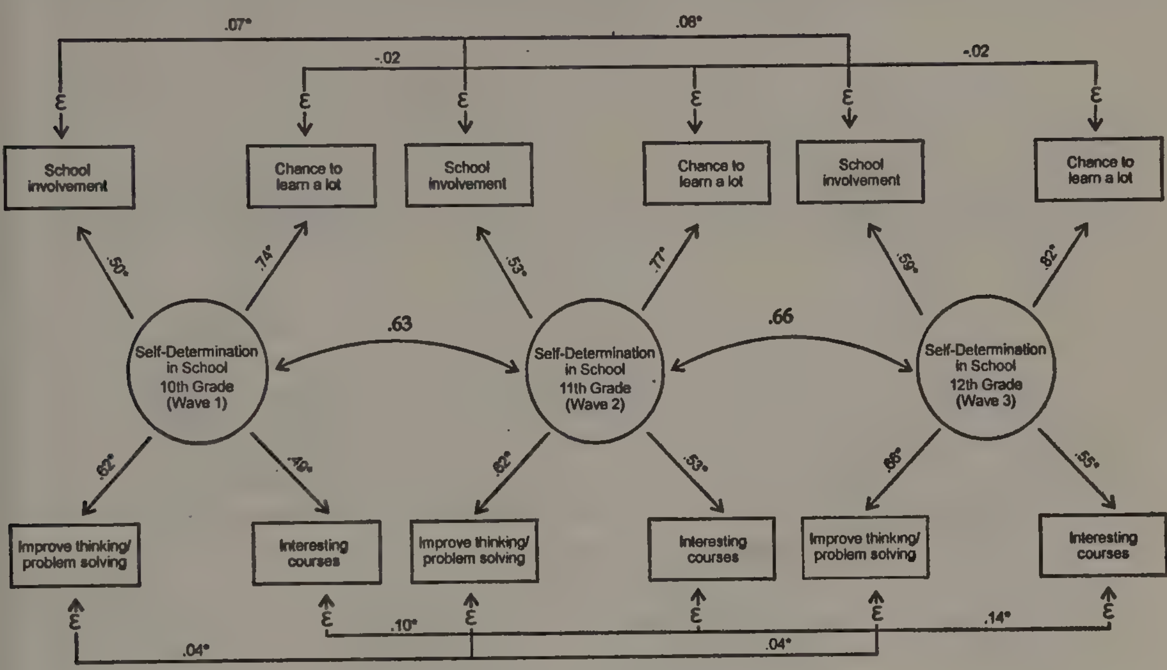
service workers, or laborers), the responses to these items were rather positive. Of the students included in this analysis, 38% said they "almost always" or "often" get a chance to plan their work with a supervisor; 46% were interested enough "most of the time" to do more work than required; and 63% felt "almost always" or "often" that their work was meaningful and important.

SELF-ESTEEM

We use four positively worded items from the Rosenberg Self-Esteem Scale (Rosenberg 1979) to form our self-image construct, which may be considered positive self-esteem or self-worth (Owens 1994). For convenience and brevity, we refer to this construct simply as self-esteem. The responses, on a 5-point scale, ranged from "almost always true" to "never true."

(1) I feel that I'm a person of worth, at least on an equal plane with others.
(2) I feel that I have a number of good qualities.
(3) I am able to do things as well as most other people.
(4) I take a positive attitude toward myself.

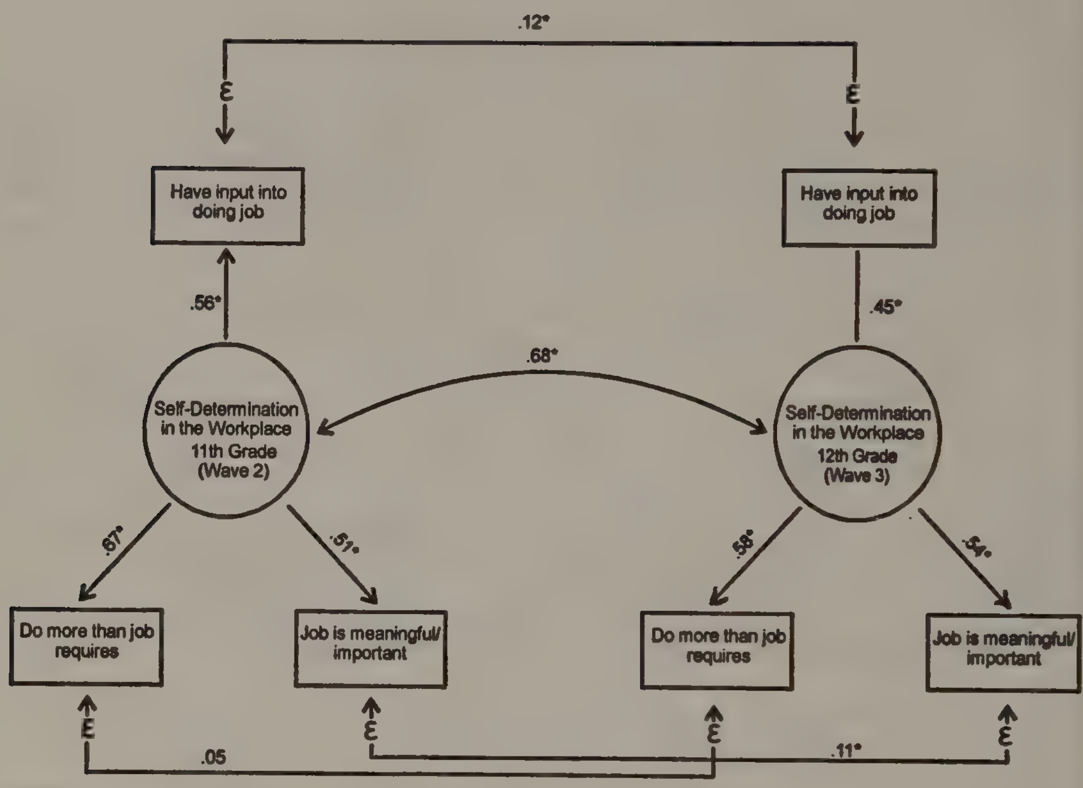
FIGURE 2: Measurement Model of Self-Determination at School



$\frac{\chi^2}{df} = \frac{202.82}{43} = 3.47$ GFI = .977 *p<.05

The theoretical and empirical bidimensionality of the self-esteem construct has been established elsewhere, with the same data set, using confirmatory factor analyses and analyses demonstrating construct validity (Owens 1993, 1994). When a unidimensional self-esteem model (including positive and negative self-esteem items in a single construct) was directly compared to a bidimensional model (consisting of two self-esteem constructs corresponding to self-depreciation and positive self-worth), the data clearly supported the bidimensional construct (Owens 1993). (Self-depreciation consists of negatively worded items such as "I feel I do not have much to proud of" and "Sometimes I think I am no good at all.") A second-order factor analysis of all positive and negative items (which posited self-depreciation and positive self-worth as derived from global self-esteem) corroborated the appropriateness of the bidimensional model (Owens 1993). Furthermore, positive self-worth was more strongly linked than self-depreciation with prosocial and proactive behaviors and attitudes and indicators of self-development; while self-depreciation was more strongly associated with depression, distress, isolation, and dependence. Since the present study's chief aim is to assess the potential *benefits* of a sense of self-determination on well-being — as expressed through one's conception of self — we focus on the positive dimension of self-esteem. (In support of our decision, analyses using a self-depreciation construct yielded little additional insight.)

FIGURE 3: Measurement Model of Self-Determination at the Workplace



$\frac{\chi^2}{df} = \frac{6.71}{5} = 1.34$ GFI = .994 *p<.05

ACHIEVEMENT

The measure of academic achievement is the respondent's self-reported grade-point average. For tenth- and eleventh-graders, queried in the fall and spring, respectively, the question read: "What is the average grade you got in your classes last year?" Twelfth-graders were asked in the spring: "What is the average grade you are getting in your classes this year?" In each case, the responses indicate the level of achievement in school over the past year. (Dornbusch, Ritter & Steinberg 1991 discuss the appropriateness of self-reported grades in comparison to grades obtained from school records; they report a correlation of .79 between these two GPA measures.)

ABILITY

The five measures of intellectual ability (Bachman 1970) are the Quick Test of Intelligence (Ammons & Ammons 1962); the Gates Test of Reading Comprehension (Gates 1958); part J, Vocabulary, and part I, Arithmetic Reasoning, of the

General Aptitude Test Battery (Super 1957); and a newly constructed test, similar to Raven's (1951) Progressive Matrices. (Their respective standardized loadings — which were derived from confirmatory factor analysis — are .70, .77, .86, .60, and .46)

SOCIOECONOMIC BACKGROUND

The socioeconomic background indicators are the father's education (.81), the mother's education (.50), and the father's occupational status — the Duncan socioeconomic index (.76).

Causal Model Specification

We estimated three models (Figures 4, 5, and 6), allowing separate examination of the influences of self-determination in the family, school, and workplace.[9] Socioeconomic status and ability are exogenous in each. Though measured at the same time as the other tenth-grade variables, they are considered prior stable phenomena. They influence tenth-grade achievement and self-esteem and the initial measures of internal motivation and control. Because of the absence of relevant data for all three time periods, there are self-determination constructs for all three periods in the school model, but for only two in the family and workplace models. The specification of direct causal influence is limited to the same or immediately contiguous time periods. For example, the effects of socioeconomic background and ability on self-esteem, measured after the tenth grade, are specified as indirect, mediated by the more proximal, relatively stable, self-image constructs.

The specification of relationships among the endogenous constructs could take a number of forms. Some statistical constraints should first be recognized. Because of multicollinearity, simultaneous estimation of both lagged and contemporaneous paths from one variable to another is hazardous. For example, if we were to specify that tenth- and eleventh-grade achievement influence eleventh-grade self-esteem, the strong relationship between these independent variables (the stability of achievement is .69 in Figure 2) could cause difficulties in parameter estimation (Gordon 1968). Second, due to the absence of appropriate instrumental variables, we do not estimate reciprocal effects.[10] As a result, in Figures 2-4 each construct is permitted to influence another only once, either through a contemporaneous or a lagged causal path, with only one-way causation.

As noted earlier, the academic achievement measures represent grades over the previous year. Intrinsic motivation and self-esteem could affect grades received subsequently, but would not influence *prior* grades. Therefore, the sense of self-determination in the family and self-esteem, both measured in the tenth grade, influence eleventh-grade achievement. However, grades received over the past year, considered as signs of competence and adaptive capacity, could influence others' granting of autonomy to the adolescent, and the adolescent's subjective perceptions of autonomy. For example, parents might allow greater independence to children who demonstrate effective problem-solving through their schoolwork. Teachers could give more challenging

assignments to students with better academic records. Past grades would also likely influence self-esteem. As a result, achievement measured in wave 1 (actually ninth-grade achievement) influences the sense of self-determination in the family and school settings in the tenth grade as well as tenth-grade self-esteem. The same rationale holds for the subsequent time period.

Of greatest interest are the interrelations of feelings of self-determination and intrinsic motivation with respect to the family, school and work, and self-esteem. These relations could be reciprocal and immediate, with intrinsic motivation influencing self-esteem and self-esteem influencing intrinsic motivation, in a dynamic ongoing process. We posit that the sense of self-determination has a contemporaneous effect on self-esteem. Assuming that perceptions of self-determination are based, at least partially, on others' actions, it is reasonable to suppose that the behaviors of persons in authority (parents, teachers, supervisors) would evolve over a period of time and be rather impervious over the short run to change in the adolescent's self-esteem. However, it is probable that over a longer period of time others' treatment would respond to change in the adolescent's self-esteem. Similarly, high self-esteem may lead one to seek out or construct complex, meaningful, and interesting activities, but it may not be possible to do so immediately.

Thus, to the extent that there is correspondence between objective situations and these subjective perceptions, we might expect lagged effects of self-image on the sense of self-determination. Prior research, using both objective and subjective measures of work experience, have generally demonstrated lagged effects of psychological dimensions on autonomy, but contemporaneous effects of autonomy on psychological variables.[11] Similarly, we posit that self-esteem has lagged effects on self-determination and immediate effects on self-esteem.

Our attention has been confined thus far to the high school period. Self-esteem one year and five years beyond high school further contribute to an understanding of self-concept development. The stability of the self-image over time indicates the degree to which earlier influences have lasting effects.

The models were estimated by maximum likelihood analysis of structural equations (Jöreskog & Sorböm 1979), using LISREL 7 (Jöreskog & Sorböm 1989). In each case, a variance-covariance matrix constituted the program input. Initially, confirmatory factor analyses were undertaken to develop the measurement models, representing each construct with items that had theoretically reasonable and substantial loadings in an exploratory factor analysis. The fact that the specified lambda-coefficients for the self-esteem and intrinsic motivation constructs remained significant and were fairly similar in magnitude across time assured us that the constructs themselves were not changing. Figures 4 to 6 are full-information models; the measurement and causal parameters are estimated simultaneously. When the same measures are repeated in multiple waves, errors in any repeated indicator may covary as a result of their shared variance or measurement effects. Thus, the models included correlated error terms of *the same* items between contiguous periods. Because we had only one indicator of grade-point average, it is assumed to be measured without error. In all figures, only the standardized parameters are presented.

Findings

THE STABILITY OF SELF-ESTEEM

The estimation and fit of our self-esteem measurement model have been extensively discussed elsewhere (Owens 1992a, 1993, 1994). Here we briefly examine the self-esteem construct across the five waves of data. First, the model fits the data well: the fit ratio is 3.43 (χ^2 = 493.65, df = 144) and the GFI is .964. The stability coefficients (or standardized betas) indicate that the construct is quite stable (see Owens 1993 for analyses showing the construct's structural invariance[12] across time). The stability of self-esteem is .62 over the first interval of the study — from the first to the second waves (18 months) — and .72 and .71 over the second and third intervals (one year each). The interval between the fourth and fifth waves increases to four years, which deflates the stability estimate to .50, yet like the other coefficients it is significant at the .05 level.

These estimates are consistent with prior theoretical and empirical work. For Erikson (1959), the establishment of identity is the critical developmental task of adolescence, after which it remains stable. The preservation of a consistent sense of self has been postulated as a major motivational goal (Epstein 1973; Korman 1970; Lecky 1945; Owens 1993; Rosenberg 1979; Swann, Stein-Seroussi & Giesler 1992). Other studies that have correlated self-image measures over time yield quite comparable estimates. Engel (1959) has estimated that self-concept stability over two-year periods during high school is .78 (corrected for unreliability). Bachman, O'Malley, and Johnston (1978), using the Youth in Transition data and a ten-item measure of global self-esteem (including our four items), presents stability estimates (corrected for unreliability) at each interval between waves of .74, .88, .89, and .69. Finally, a panel study of male students indicated a similar level of persistence for four self-concept dimensions during college (Mortimer & Lorence 1981). These findings indicate that as a construct, self-image tends to be quite stable (from adolescence onward); and, with the possible exception of a slight dip in the youngest adult years, moves on a trajectory of increased stability over the life span (see Alwin 1994).

We turn now to three causal models corresponding to the family, the school, and the workplace. Each model appears to fit the data well. In the family model (Figure 4) the fit ratio is 2.42 (χ^2 = 1595.24, df = 659) and the GFI is .921. In the school model (Figure 5), the fit ratio is 2.49 (χ^2 = 2015.12, df = 809) and the GFI is .911. Finally, the fit ratio in the work model (Figure 6) is again reasonable: 1.56 (χ^2 = 917.63, df = 587) with a GFI of .828.

THE FAMILY MODEL

It is noteworthy that the construct representing self-determination with respect to the family (Figure 4) is quite persistent (.65). (All asterisked coefficients are statistically significant, $p < .05$.) Clearly, parents' treatment of the adolescent, at least as seen through the eyes of the child, does not change radically from year to year.

Social class background has a positive effect on the perception of internal control in the family (.16), indicating a more democratic style of relationships in higher-status families (Gecas 1979). While the findings suggest that parents do

not confer greater decision-making power on children of higher intellectual ability, demonstrated achievement does matter: school achievement prior to the tenth grade has a positive effect on internal control in the family sphere (.19) in the tenth grade. By the eleventh grade, however, neither academic achievement nor the self-esteem of the child has an observable effect on the sense of self-determination, net of its stability over time.

The paths of greatest interest are those linking the sense of self-determination in the family sphere and self-esteem. The coefficient from the tenth-grade construct to self-esteem suggests that this effect is rather strong (.38), with grade-point average and ability controlled. However, because the stability of self-esteem over time is not taken into account here, the second path — from self-determination in the family to self-esteem in the eleventh grade — offers a more crucial test. This path of .13 may not seem particularly large, *but both grades and prior self-esteem are now controlled.* Adolescent experiences of internal control in the family domain thus appear to enhance self-esteem. Given the high stability of self-esteem (.79, .75, and .54 over the next three waves), this influence, though small in absolute magnitude, is likely to be persistent (Kohn & Schooler 1978).

In the mid portion of the model, the path from ability to student grade-point average in wave 1 (tenth grade) is a rather striking .62. At the same time, the path from socioeconomic background is a negative .10. It would appear that teachers, through their assignment of grades, reward ability, not social class origin (Rehberg & Rosenthal 1978; Rosenberg, Schooler & Schoenbach 1989). It should be noted, however, that the multicollinearity (.42) of the predictors could generate instability in the beta-coefficients. Whereas the zero-order correlation of the socioeconomic background and achievement (GPA) constructs is a positive .178, other research has also reported negligible relations between school grades and social class background once ability is controlled (Crouse et al. 1979; Portes & Wilson 1976; Sewell & Hauser 1975). The only significant predictor of grades in the eleventh grade is the lagged variable, and grades exhibit high stability. In the twelfth grade, self-esteem has a positive effect on grades (.08), which comports with Owens (1994).

The fact that socioeconomic background has no significant effect on self-esteem in grade ten is consistent with Rosenberg's and Pearlin's (1978) claim that it is one's own achievement, not that of one's parents, that determines self-evaluation. Also consistent with their position, both measured ability (.15) and grade-point average over the previous year (.11) have positive effects on tenth-grade self-esteem. Grade-point average through the eleventh grade (measured in wave 2) likewise influences self-esteem in the spring of grade eleven (.17). (Owens 1994 and Spenner and Otto 1985 also find that grades have positive effects on self-esteem.) However, by the end of high school, grades no longer affect the adolescent self-image. Bachman, O'Malley, and Johnston (1978) have reported the declining influence of grades over time; they attribute this pattern to a developmental effect, not to secular decline in the value of education over the period (O'Malley & Bachman 1979).

School performance (indicated here by GPA) may have little impact on twelfth-grade self-esteem (unlike that of sophomores and juniors) because of the expanded freedom and widening social activities and social comparisons

available to a graduating senior (e.g., dating, driving, working, college-hunting, preparing to enter the military, and preparing to move out on one's own). As a result, the earlier importance of high school grades as a source of self-esteem is diluted. Moreover, by the twelfth grade, the asset or liability of grades as a source of post-high-school opportunities (Owens 1992b) has, for the most part, already been felt, distancing the psychological centrality afforded high school grades.

THE SCHOOL MODEL

As Figure 5 shows, there is also a rather high stability in the intrinsic motivation to schoolwork over the high school period (both betas = .62). The consistency of school achievement (GPA) as an influence on the perception of self-directed activity in school is particularly striking. School grades in the first (.29), second (.12), and third (.14) waves foster an internal motivation for schoolwork. Apparently, students with high grades come to find schoolwork more meaningful and interesting and feel they are improving their ability to think and solve problems. At the same time, intrinsic motivation at school influences grades, as indicated by the paths of .08 and .11 over the first and second intervals.

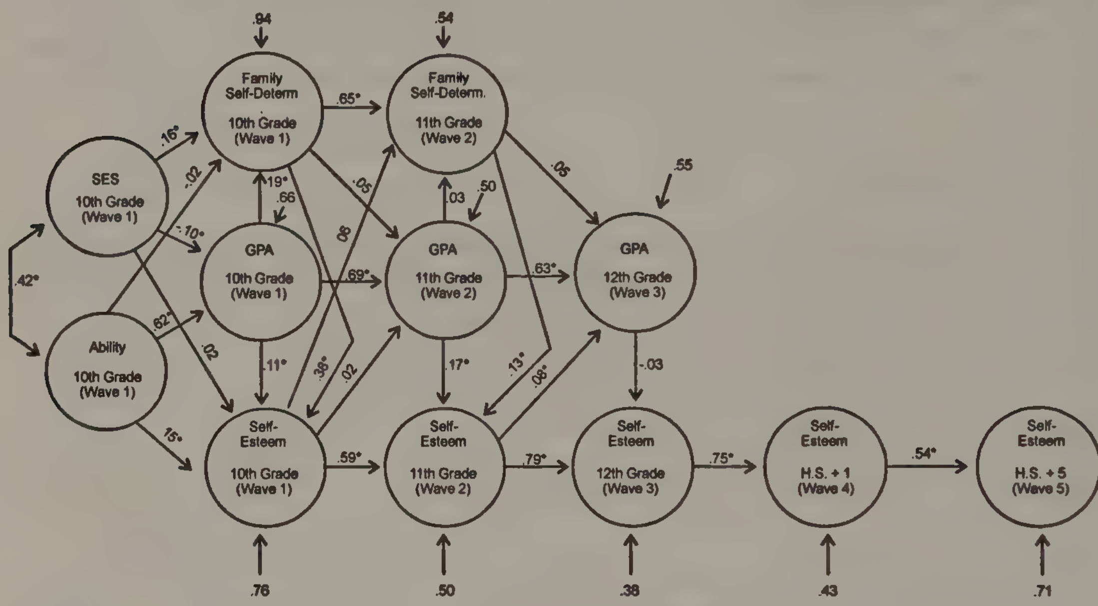
While there was little evidence that self-determination in the family setting and grades are positively interrelated, here we find a pattern of mutual influence. One might expect that a sense of intrinsic motivation and control in the domain in which grades are achieved, that of the school, would have greater significance for achievement. Unlike the family construct, intrinsic motivation for schoolwork is not affected by parental social class.[13] Moreover, student ability has no significant independent effect on intrinsic motivation in school. To the extent that intrinsic motivation is an indicator of students' experiences in the school (not simply reflections of social class and ability), it may be concluded that teachers, in promoting these experiences, respond to students' actual performances.

What are most important from our perspective are the three significant paths from the intrinsic motivation construct to self-esteem. They are manifest with ability and grades taken into account in the first period (.28), and when grades and prior self-esteem are controlled in the second (.14) and third (.16) periods. Apparently, it is not so much the achievement of high grades that enhances self-esteem; the experience of intrinsic motivation in school generally has the more consistent effect. A sense of self-initiated action and intrinsic motivation enhance the self. The fact that self-esteem has no significant effects on intrinsic motivation in school, in either interval, argues against the notion that our measure of intrinsic motivation is simply a reflection of self-esteem.[14]

THE WORK MODEL

Before describing the findings on work (Figure 6), we remind the reader that this analysis includes only students employed during *both* the eleventh and the twelfth grades. (The measures of intrinsic motivation for work were not included in the tenth-grade survey.) Intrinsic motivation for work manifests somewhat higher stability (.76) than the corresponding family and school constructs in the whole panel. However, school grades are somewhat less stable

FIGURE 4: Family Self-Determination and Self-Esteem



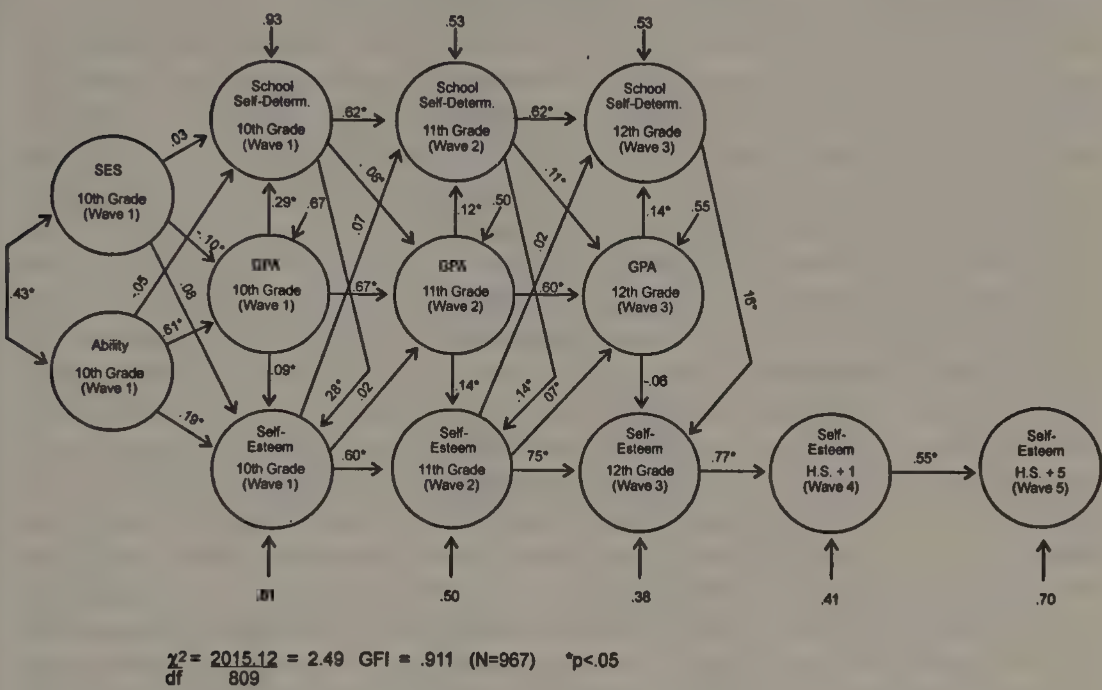
$\frac{\chi^2}{df} = \frac{1595.24}{659} = 2.42$ GFI = .921 (N=965) *p<.05

in this subgroup between the eleventh and twelfth grades (.50). And self-esteem also shows somewhat less stability through the high school period (but not thereafter), especially in the path from tenth- to eleventh-grade self-esteem (.46). These coefficients express the degree of correlational stability — the extent to which persons who have high scores during the first period, relative to their peers, retain this advantage during the second.[15]

While the differences in stability are not large, they suggest that students who continue to work while in school may experience more change.[16] Social scientists studying the life span have emphasized the importance of environmental continuity as a source of personal stability (Mortimer, Finch & Kumka 1982; Moss & Sussman 1980; Owens 1992a; Sears 1981). According to "the aging stability hypothesis" (Glenn 1980), younger people are more likely to have to adapt to new roles and social conditions; their personalities therefore are less stable (Lorence & Mortimer 1985; Owens 1992a). The adolescents in this working subgroup have recently encountered a new life domain, probably necessitating major adaptations in daily routines. The somewhat lower stability of this group in self-esteem and achievement might be expected given this new role acquisition.

The path from tenth-grade self-esteem to eleventh-grade intrinsic work motivation is positive (.33), while eleventh-grade self-esteem bears a *negative* relation to the twelfth-grade work construct (-.22). Recent developments in a bidimensional view of self-esteem as well as aspects of the principle of self-actualization help explain this seeming paradox. First, if Owens (1993, 1994) is

FIGURE 5: School Self-Determination and Self-Esteem



correct in asserting that the positive dimension of self-esteem is associated with prosocial and proactive attitudes and behaviors, then individuals with high self-esteem entering the world of work may seek — and be granted — responsibilities and input into the job unavailable to those with less favorable self-images. The result is that positive self-esteem initially increases self-determination at work. Second, as one continues on the job — as all the subjects in this subsample did — initial intrinsic motivations may subside as the constraints of adolescent work routines become manifest (nearly three-fourths of the workers were operatives, service workers, or laborers). Consequently, as one stays in the high school labor force, greater self-esteem may actually work against *subjective* feelings of self-determination on the job, thwarting high-self-esteem subjects' attempts to experience self-actualization (Maslow 1970). Finally, a form of dissonance may set in as the high-self-esteem individual — previously excited about the prospects of pitching in and having a say — exhausts the intrinsic elements and growth potential of the job, culminating in a new feeling that one's assets are not being fully utilized.[17]

We observed earlier that grades in school predicted intrinsic motivation, in the case of the family in the first period and of the school in all three periods. Here we find that achievement in school has no significant effect on intrinsic motivation at work. Moreover, an intrinsic motivation toward work does not enhance school grades. Intrinsic motivation in the work sphere, unlike that in the school, appears to be quite unrelated to school achievement. (Other analyses of these data suggest that this subjective experience of work, in conjunction with long work hours, depresses school achievement — see Finch & Mortimer 1985.)

Still, grades in school show a positive relation to self-esteem among employed students in both the tenth (.28) and eleventh (.20) grades.

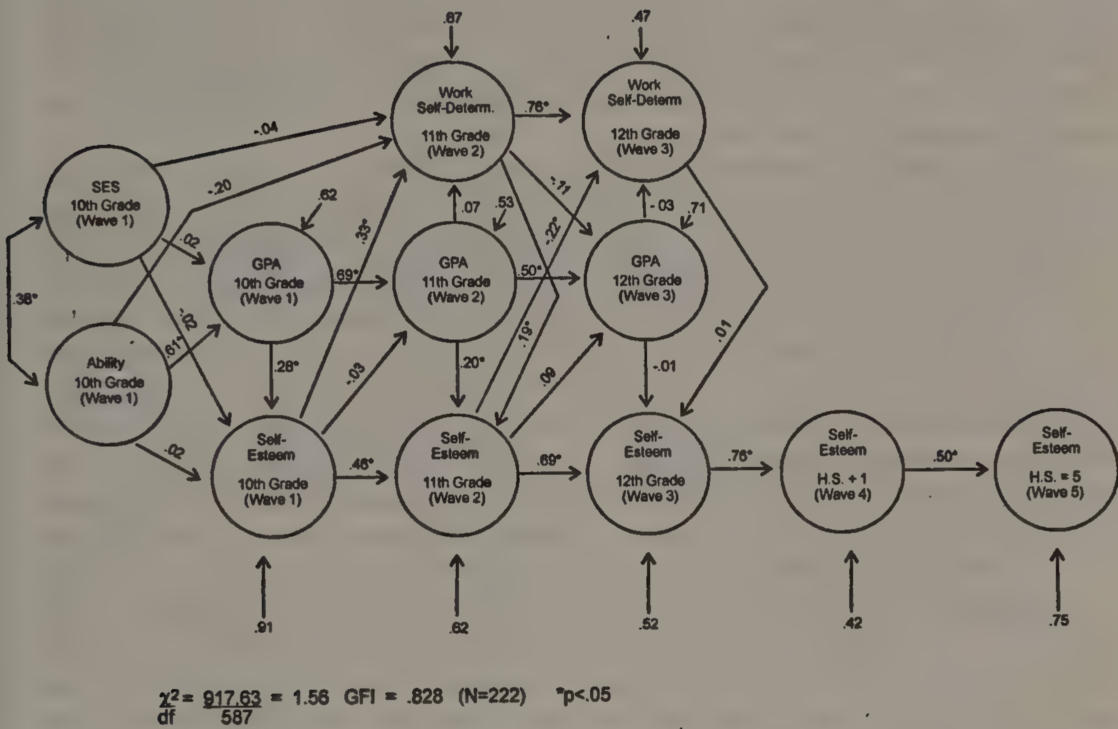
Turning finally to the paths of greatest pertinence, we observe a perplexing pattern. Intrinsic motivation toward work enhances self-esteem in the eleventh grade (.19), confirming the hypothesis that this subjective experience on the job engenders positive self-attributions and evaluations. But intrinsic motivation with respect to work is not significantly related to self-esteem in the final year in high school. Persons may be most vulnerable to change — of a positive or negative character — immediately after the acquisition of new social roles (Frese 1982; Hall 1971; Owens 1992a; Van Maanen & Schein 1979). We selected this subgroup on the basis of their work experience over two contiguous years. It is conceivable that by the second year, the immediate impact of work experience on self-esteem would have dissipated (the effect of working would then become wholly indirect — through the stability of self-esteem). To examine the evidence for this interpretation, we compared the effects of intrinsic motivation at work on self-esteem (using ordinary least-squares regression) for those who started work in the eleventh grade and those who were working in both the tenth and eleventh grades. For the first group, "new" to employment in the junior year, intrinsic work motivation had a significant effect on eleventh-grade self-esteem ($\beta = .17$, $p < .001$, with achievement and tenth-grade self-esteem controlled), while intrinsic work motivation was not significantly related to self-esteem for the veteran workers ($\beta = .08$). However, in the senior year the effects of intrinsic work motivation on self-esteem were not statistically significant for either the new workers (those working for the first time in grade 12) or the veteran workers (those who also worked in grade 11). We conclude that intrinsic motivation in the work sphere, like that in the family and school domains, can have positive implications for the self. In this panel, however, the positive effect is confined to the junior year of high school.[18]

Discussion

The limitations of this research should be remedied in the future. Regrettably, this study does not include both women and men; it is still necessary to determine whether intrinsic motivation and a sense of internal control have similar implications for self-concept development in adolescent women. The work of Miller et al. (1979; Miller 1980) suggests that adult men and women respond similarly to self-direction in work (see also Mortimer, Finch & Maruyama 1988).

A second limitation is that all constructs, with the exception of ability, are self-reported. For self-esteem, whose essence is truly subjective, such measures are optimal. Our measures of self-determination are also subjective. It is of interest to know to what extent the subjective perception of self-directed activity, what we call self-determination or intrinsic motivation, corresponds to actual opportunities for self-directed activities in the three role domains under investigation. To assess the amount of autonomy inherent in the environment, information from parents, teachers, and supervisors or other more objective data would be highly desirable.

FIGURE 6: Work Self-Determination and Self-Esteem



Third, in an attempt to simplify the causal models, we have not simultaneously assessed the effects of intrinsic motivation in the family, school, and workplace on the development of self-esteem, in the context of the same analytic model, nor have we examined the contemporary reciprocal interrelations of the sense of self-determination and self-esteem. The independent effects of experiences in these domains on the self-concept, and any immediate reciprocal causality, remain to be determined.

But in spite of these limitations, these data are consistent with the proposition that the sense of self-determination in three domains of life enhances adolescent self-esteem. Supporting Kohn and Schooler's "generalization model," and Rosenberg's "principle of self-attribution," shared decision making in the family and an internal locus of motivation in school engender positive self-evaluations. The findings were mixed with respect to the work sphere. This suggests that the intrinsic motivation toward work may have somewhat different developmental implications for adolescents (see also Finch & Mortimer 1985; Mortimer & Finch 1986a). However, longevity in teenage employment may counteract perceptions of self-determination. Nevertheless, considering all three context domains, the results of this research generally parallel a previous study of a panel of young, highly educated adult men (Mortimer & Lorence 1979b; Mortimer, Lorence & Kumka 1986) for whom autonomy was found to strengthen the sense of competence over time. Moreover, the findings confirm and extend Kohn and his colleagues' studies of a more representative panel, including the full range of age groups (16 and over) in the working population

(Kohn & Schooler 1982; Slomczynski, Miller & Kohn 1981), and both sexes (Miller et al. 1979).

There was evidence that subjective perceptions of self-direction in school, but not in the workplace, also promote academic achievement. Perceiving one's schoolwork as providing opportunities to learn, think, and solve problems consistently heightened school grades. Prior school grades also enhanced intrinsic motivation at school.

This article has focused on the interrelations of the sense of self-determination, school achievement, and self-esteem. The finding that each of these constructs is rather stable has important implications. The stability of the self-determination constructs indicates considerable continuity in adolescent experiences in the family, at school, and at work. The stability of achievement and self-esteem suggests that we may have to look elsewhere, to early adolescence or even to childhood, to identify the most crucial formative experiences.

Most past studies have examined socialization and development in each life stage as a distinct process; there are literatures on infancy, childhood, adolescence, adulthood, and old age (Mortimer & Simmons 1978). The present research suggests the fruitfulness of an alternative perspective on the life course. The same kinds of experiences that have been found to stimulate positive psychological outcomes in adulthood — personal discretion in decision making, a sense of self-determination, interesting and challenging tasks — also foster self-esteem in the adolescent years. But studies of adults are generally confined to the domain of work. Here we find that the sense of self-determination, a probable subjective outcome of autonomy, has consequences for the self when it is encountered in other social contexts, in the family and at school, as well as in the workplace. The findings suggest that autonomy has important implications for human development (Bronfenbrenner 1979), though it may be expressed in different ways in different periods and domains of life. They indicate that there are gains in considering the life course as a whole, so that the insights developed from research on one phase can enlighten investigations of others.

Notes

1. In a broader sense, Woelfel and Haller (1971) argue that the role of the "self-reflective act" in attitude formation (including attitudes about the self) is more potent than interpersonal influences when the "object of the attitude is unambiguous and observable" (76).

2. We can place our investigation of self-determination's relation to self-esteem in the broader literature on the relation of perceptions of self-control to general well-being (e.g., social, psychological, and physical). For example, using a longitudinal sample of Depression-era youths, Clausen (1993) finds that planfully competent adolescent males, who also have a higher sense of personal control and self-determination, have greater psychological and physical well-being, smoother interpersonal relationships, fewer recurrent life crises, and less failure and frustration in later life than their less competent peers. Such related perceived control concepts as self-efficacy (Bandura 1989), locus of control (Strickland 1989), and personal control (Rodin, Timko & Harris 1985) have been linked to such important outcomes as health, sports, school, work, marriage, and general life satisfaction (also see Skinner 1995).

3. It has also been shown that perceptions of greater control in the classroom lead to a variety of positive school outcomes, such as increased effort, attention, and participation, as well as improved self-regulatory strategies (Berry & West 1993; Skinner 1995).

4. According to Condry (1977:472), "the outcomes of intrinsically motivated activity are encoded as more self-relevant and thus internalized more than the outcomes of extrinsically motivated activity, where the role of the self is more passive."

5. Overall cumulative attrition by wave was 14.8% for wave 2, 18.7% for wave 3, 26.8% for wave 4, and 26.4% for wave 5. Attrition comparisons between the 291 blacks and the 2,177 whites are wave 2, 13.4% for whites and 17.5% for blacks; wave 3, 17.9% for whites and 18.6% for blacks; wave 4, 25.3% for whites and 30.2% for blacks; and wave 5, 29.5% for whites and 46.4% for blacks.

6. Excluding persons with missing data led to the loss of 627 cases (37% of the 1,681 white respondents in the fifth wave). To ascertain the differences between them and the other respondents, a series of t-tests was performed, using additive construct scores. There were no significant differences between persons with no missing data and those with missing data (on any other variable) in socioeconomic background, family autonomy at time 2, or self-esteem after the tenth grade. Respondents with missing data were lower on ability; on family autonomy at time 1; on school autonomy at times 1 and 2; on achievement at times 1, 2, and 3; and on self-esteem in tenth grade. These differences are cause for concern. But pairwise deletion also introduces an indeterminate amount of error, because the entries in a data matrix are not based on the same cases. Moreover, covariance matrices based on incomplete data are often ill-defined and occasionally not positive definite. The most obvious problem is choosing the appropriate standard deviations (SD). In computing the correlation or the covariance of the variables j and k, when some data are missing it is not the case that SD_j is constant for all pairs (SD_j, SD_k). While various proposals have been advanced to alleviate this problem, we conservatively limit the analysis to cases without missing data (Bollen 1989).

7. We report the χ^2/degrees of freedom ratio (fit ratio) and the goodness-of-fit index (GFI). The fit ratio generally should not exceed about 3 to 5. A nonsignificant fit ratio indicates that the variance-covariance matrix derived from the theoretical model is similar to the observed variance-covariance matrix, as one would hope. The GFI shows the amount of variance and covariance explained by the model, and ranges from 0 (no account) to 1 (perfect account). A value in the range of .9 is generally acceptable, although there is no clear consensus on the cutoff (Jöreskog & Sorböm 1989).

8. Deci, Eghrari, and Patrick (1994) suggest that self-determination and intrinsic motivation are essentially equivalent concepts, giving us confidence that our measure of self-determination in the school is appropriate.

9. Since the self-determination constructs involve three separate domains and different actors (parents, teachers, and supervisors), they are conceptually distinct. Empirically, they are only moderately intercorrelated. (In grade ten, the correlation between self-determination in the family and at school, using unit weighted indices, is .30; in grade eleven, it is .27 between family and work and .25 between school and work.) An exploratory analysis (principal factor) including the family and school items yielded two distinct factors corresponding to each domain in the tenth and eleventh grades.

10. Because we do not include reciprocal effects, our analyses could overestimate the one-way effects.

11. We believe that the models are justified on conceptual grounds (Mortimer & Lorence 1979a, 1979b) and on the basis of prior empirical work (Kohn & Schooler 1978, 1982, 1983; Lorence & Mortimer 1985; Mortimer & Finch 1986a). Still, an alternative causal ordering of the variables cannot be ruled out.

12. Structural invariance is present when the dimensions of a construct are the same across time and when the pattern of relationships among its measured attributes persists across waves.

13. In contrast, Miller, Kohn, and Schooler (1982) report that parental socioeconomic status has a positive effect on educational self-direction, as indicated by the substantive complexity of schoolwork and closeness of teacher supervision.

14. A replication of this model using an alternative self-depreciation construct shows that GPA has no causal effect on self-depreciation, while self-depreciation engenders lower grades (-.08 and -.09 for the paths from tenth to eleventh and from eleventh to twelfth grade, respectively). Exploring the alternative self-depreciation models further, we note that the relation of wave 1

self-determination in the family to wave 2 self-depreciation = -.35 ($p < .05$) and of wave 1 self-depreciation to wave 2 self-determination in the family = -.08. Self-depreciation's relationship to school self-determination is wave 1 school to wave 2 self-depreciation = -.32 ($p < .05$), wave 1 self-depreciation to wave 2 school = -.02; wave 2 family to wave 2 self-depreciation = -.13 ($p < .05$), and wave 2 self-depreciation to wave 3 family = -.01. For the workplace model the relationships are wave 1 self-depreciation to wave 2 work = -.39 ($p < .05$); wave 2 work to wave 2 self-depreciation = -.09; wave 2 self-depreciation to wave 3 work = -.07.

15. Another way of conceptualizing stability is by level change (Mortimer, Finch & Kumka 1982). Using this criterion, both employed and unemployed students increased in self-esteem over time.

16. However, when the family and school models were reestimated for this working subgroup, intrinsic motivation in both family and school manifested levels of stability that were almost identical to those in the whole panel.

17. Another paradox occurs when we see some evidence for the first time that ability and an intrinsic motivation construct may be negatively correlated ($\gamma = -.20$). (Although the t-value of 1.9 is just below the conventional cutoff of $t = 1.96$ for $p < .05$, given that this subsample's N of 222 is less than one-quarter the size of the other two models, this gamma-coefficient is interesting.) Why would the less able students feel more internally motivated in their jobs? The picture becomes clearer when we consider the difference in time working by student ability. Students of lower ability were working longer hours (the correlation of ability to hours is a negative .28 in the eleventh grade). Given their greater time involvement in the work sphere, the less able students may actually have the more responsible jobs. On the average, the eleventh- and twelfth-grade students (who were employed in both the junior and senior years) were working approximately twenty-four hours per week.

18. Shanahan et al. (1991) found a positive effect of decision-making autonomy on high school students' depressive affect, indicating that novice workers find independence somewhat unnerving.

References

Alwin, Duane F. 1994. "Aging, Personality and Social Change." Pp. 135-85 in *Life-Span Development and Behavior*. Vol. 12, edited by David L. Featherman, Richard M. Lerner, and Marion Perlmutter. Lawrence Erlbaum.

Ammons, R.B., and C.H. Ammons. 1962. "The Quick Test (QT): Provisional Manual." *Psychological Reports* 11 (Monograph suppl. 7,7).

Andrisani, Paul J., and Gilbert Nestel. 1976. "Internal-External Control as Contributor to and Outcome of Work Experience." *Journal of Applied Psychology* 61:156-65.

Bachman, Jerald G. 1970. *Youth in Transition*. Vol. 2, *The Impact of Family Background and Intelligence on Tenth-Grade Boys*. Survey Research Center, Institute for Social Research, University of Michigan.

Bachman, Jerald G. 1974. *Young Men in High School and Beyond: A Summary of Findings of the Youth in Transition Project, 1966-74* [MRDF]. Survey Research Center, University of Michigan [producer]. Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor].

Bachman, Jerald G., Patrick M. O'Malley, and Jerome Johnston. 1978. *Youth in Transition*. Vol. 6, *Adolescence to Adulthood — Change and Stability in the Lives of Young Men*. Survey Research Center, Institute for Social Research, University of Michigan.

Bandura, Albert. 1989. "Human Agency in Social Cognitive Theory." *American Psychologist* 44:1175-84.

Barber, Brian K., and Boyd C. Rollins (eds.). 1990. *Parent-Adolescent Relationships*. University Press of America.

Bem, Daryl J. 1972. "Self-Perception Theory." Pp. 1-62 in *Advances in Experimental Social Psychology*. Vol. 6., edited by Leonard Berkowitz. Academic Press.

Berry, Jane M., and Robin L. West. 1993. "Cognitive Self-Efficacy in Relation to Personal Mastery and Goal Setting across the Life Span." *International Journal of Behavioral Development* 16:351-79.

Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. John Wiley & Sons.

Bowerman, Charles E., and Glen H. Elder Jr. 1964. "Variations in Adolescent Perception of Family Power Structure." *American Sociological Review* 29:551-67.

Bronfenbrenner, Urie. 1979. *The Ecology of Human Development*. Harvard University Press.

Brookover, Wilbur B., Shailer Thomas, and Ann Paterson. 1964. "Self-Concept of Ability and School Achievement." *Sociology of Education* 37:271-78.

Calsyn, Robert J., and David A. Kenny. 1977. "Self-Concept of Ability and Perceived Evaluation of Others: Cause or Effect of Academic Achievement?" *Journal of Educational Psychology* 69:136-45.

Clausen, John A. 1991. "Adolescent Competence and the Shaping of the Life Course." *American Journal of Sociology* 96:805-42.

____. 1993. *American Lives: Looking Back at the Children of the Great Depression*. Free Press.

Cohn, Richard M. 1978. "The Effect of Employment Status Change on Self-Attitudes." *Social Psychology* 41:81-93.

Coleman, James S. 1974. *Youth: Transition to Adulthood*. University of Chicago Press.

Condry, John. 1977. "Enemies of Exploration: Self-Initiated versus Other-Initiated Learning." *Journal of Personality and Social Psychology* 35:459-77.

Cooley, Charles Horton. 1902. *Human Nature and the Social Order*. Scribner's.

Covington, Martin V. 1992. *Making the Grade: A Self-Worth Perspective on Motivation and School Reform*. Cambridge University Press.

Crouse, James, Peter Meuser, Christopher Jencks, and Charles S. Reichardt. 1979. "Latent Variable Models of Status Attainment." *Social Science Research* 8:348-68.

Deci, Edward L. 1975. *Intrinsic Motivation*. Plenum.

Deci, Edward L., Haleh Eghrari, and Brian C. Patrick. 1994. "Facilitating Internalization: The Self-Determination Theory Perspective." *Journal of Personality* 62:119-42.

Deci, Edward L., and Richard M. Ryan. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. Plenum.

Dornbusch, Sanford M. 1989. "The Sociology of Adolescence." *Annual Review of Sociology* 15:233-59.

Dornbusch, Sanford M., Philip L. Ritter, and Laurence Steinberg. 1991. "Community Influences on the Relation of Family Statuses to Adolescent School Performance: Differences between African Americans and Non-Hispanic Whites." *American Journal of Education* 99:543-67.

Elder, Glen H. Jr. 1962. "Structural Variations in the Child Rearing Relationship." *Sociometry* 25:241-62.

____. 1963. "Parental Power Legitimation and Its Effects on the Adolescent." *Sociometry* 26:50-65.

____. 1968. *Adolescent Socialization and Personality Development*. Rand McNally.

____. 1969. "Occupational Mobility, Life Patterns, and Personality." *Journal of Health and Social Behavior* 10:308-23.

____. 1974. *Children of the Great Depression*. University of Chicago Press.

Engel, Mary. 1959. "The Stability of the Self-Concept in Adolescence." *Journal of Abnormal and Social Psychology* 58:211-15.

Epstein, Seymour. 1973. "The Self-Concept Revisited: Or a Theory of a Theory." *American Psychologist* 28:404-16.

Erikson, Erik H. 1959. "The Problem of Ego Identity." *Psychological Issues* 1:101-64.

Finch, Michael D., and Jeylan T. Mortimer. 1985. "Adolescent Work Hours and the Process of Achievement." Pp. 171-96 in *Research in Sociology of Education and Socialization*. Vol. 5, edited by Alan C. Kerckhoff. JAI Press.

Finch, Michael D., Michael J. Shanahan, Jeylan T. Mortimer, and Seongryeol Ryu. 1991. "Work Experience and Control Orientation in Adolescence." *American Sociological Review* 56:597-611.

Franks, David D., and Joseph Marolla. 1976. "Efficacious Action and Social Approval as Interacting Dimensions of Self-Esteem: A Tentative Formulation through Construct Validation." *Sociometry* 39:324-41.

Frese, Michael. 1982. "Occupational Socialization and Psychological Development." *Journal of Occupational Psychology* 55:209-24.

Gates, Arthur Irving. 1958. *Gates Reading Survey — Form 1*. Teachers College, Columbia University.

Gecas, Viktor. 1979. "The Influence of Social Class on Socialization." Pp. 365-404 in *Contemporary Theories about the Family*. Vol. 1, *Research-Based Theories*, edited by Wesley R. Burr, Reuben Hill, F. Ivan Nye, and Ira L. Reiss. Free Press.

Gecas, Viktor, and Michael L. Schwalbe. 1983. "Beyond the Looking-glass Self: Social Structure and Efficacy-Based Self-Esteem." *Social Psychology Quarterly* 46:77-88.

Glenn, Norval D. 1980. "Values, Attitudes, and Beliefs." Pp. 596-640 in *Constancy and Change in Human Development*, edited by Orville G. Brim Jr. and Jerome Kagan. Harvard University Press.

Goodwin, Leonard. 1977. *The Work Incentive (WIN) Program and Related Experiences*. U.S. Government Printing Office.

Gordon, Robert A. 1968. "Issues in Multiple Regression." *American Journal of Sociology* 73:592-616.

Greenberger, Ellen, and Laurence Steinberg. 1986. *When Teenagers Work: The Psychological and Social Costs of Adolescent Employment*. Basic Books.

Greenberger, Ellen, Laurence D. Steinberg, and Mary Ruggiero. 1982. "A Job Is a Job Is a Job . . . Or Is It?" *Work and Occupations* 9:79-96.

Griffin, Ricky W. 1983. "Objective and Social Sources of Information in Task Redesign: A Field Experiment." *Administrative Science Quarterly* 28:184-200.

Gurin, Gerald, and Patricia Gurin. 1976. "Personal Efficacy and the Ideology of Individual Responsibility." Pp. 131-57 in *Economic Means for Human Needs*, edited by Burkhard Stumpel. Survey Research Center, Institute for Social Research, University of Michigan.

Hall, Douglas T. 1971. "A Theoretical Model of Career Subidentity Development in Organizational Settings." *Organizational Behavior and Human Performance* 6:50-76.

House, James S. 1980. *Occupational Stress and the Mental and Physical Health of Factory Workers*. Survey Research Center, Institute for Social Research, University of Michigan.

———. 1981. "Social Structure and Personality." Pp. 525-61 in *Social Psychology*, edited by Morris Rosenberg and Ralph H. Turner. Basic Books.

Jacques, Jeffrey M., and Karen J. Chason. 1977. "Self-Esteem and Low Status Groups: A Changing Scene?" *Sociological Quarterly* 18:399-412.

James, William. 1892. *Psychology*. Henry Holt.

Jöreskog, Karl G., and Dag Sorböm. 1979. *Advances in Factor Analysis and Structural Equation Models*. Abt.

———. 1989. *LISREL 7: A Guide to the Program and Applications*. 2d ed. SPSS.

Kandel, Denise B., and Gerald S. Lesser. 1972. *Youth in Two Worlds*. Jossey-Bass.

Kelley, Harold H., and John L. Michela. 1980. "Attribution Theory and Research." *Annual Review of Psychology* 31:457-501.

Kohn, Melvin L. 1969. *Class and Conformity: A Study in Values.* Dorsey.

____. 1976. "Looking Back — A 25 Year Review and Appraisal of Social Problems Research." *Social Problems* 24:94-112.

____. 1980. "Job Complexity and Adult Personality." Pp. 193-210 in *Themes of Work and Love in Adulthood,* edited by Neil J. Smelser and Erik H. Erikson. Harvard University Press.

____. 1981. "Personality, Occupation, and Social Stratification: A Frame of Reference." Pp. 267-97 in *Research in Social Stratification and Mobility.* Vol. 1, edited by Donald J. Treiman and Robert V. Robinson. JAI Press.

Kohn, Melvin L., and Carmi Schooler. 1973. "Occupational Experience and Psychological Functioning: An Assessment of Reciprocal Effects." *American Sociological Review* 38:97-118.

____. 1978. "The Reciprocal Effects of the Substantive Complexity of Work and Intellectual Flexibility: A Longitudinal Assessment." *American Journal of Sociology* 84:24-52.

____. 1982. "Job Conditions and Personality: A Longitudinal Assessment of Their Reciprocal Effects." *American Journal of Sociology* 87:1257-86.

____. 1983. *Work and Personality: An Inquiry into the Impact of Social Stratification.* Ablex.

Korman, Abraham K. 1970. "Toward an Hypothesis of Work Behavior." *Journal of Applied Psychology* 54:31-41.

Lamborn, Susie D., Nina S. Mounts, Laurence Steinberg, and Sanford M. Dornbusch. 1991. "Patterns of Competence and Adjustment among Adolescents from Authoritative, Authoritarian, Indulgent, and Neglectful Families." *Child Development* 62:1049-65.

Lecky, Prescott. 1945. *Self-Consistency: A Theory of Personality.* Island Press.

Lorence, Jon, and Jeylan T. Mortimer. 1985. "Job Involvement through the Life Course: A Panel Study of Three Age Groups." *American Sociological Review* 50:618-38.

Luck, Patrick W., and Jerold Heiss. 1972. "Social Determinants of Self-Esteem in Adult Males." *Sociology and Social Research* 57:69-84.

Marcus, Hazel, Susan Cross, and Elissa Wurf. 1990. "The Role of the Self-System in Competence." Pp. 205-25 in *Competence Considered,* edited by Robert J. Sternberg and John Kolligian Jr. Yale University Press.

Maruyama, Geoffrey, Michael D. Finch, and Jeylan T. Mortimer. 1985. "Processes of Achievement in the Transition to Adulthood." Pp. 61-87 in *Current Perspectives on Aging and the Life Cycle,* edited by Zena Smith Blau. JAI Press.

Maslow, Abraham H. 1970. *Motivation and Personality.* 2d ed. Harper & Row.

McLeod, Jane D., and Michael J. Shanahan. 1993. "Poverty, Parenting, and Children's Mental Health." *American Sociological Review* 58:351-66.

Mead, George H. 1934. *Mind, Self, and Society,* edited and introduction by Charles W. Morris. University of Chicago Press.

Miller, Joanne. 1980. "Individual and Occupational Determinants of Job Satisfaction: A Focus on Gender Differences." *Sociology of Work and Occupations* 7:337-66.

Miller, Joanne, Carmi Schooler, Melvin L. Kohn, and Karen Miller. 1979. "Women and Work: The Psychological Effects of Occupational Conditions." *American Journal of Sociology* 85:66-94.

Miller, Karen A., Melvin L. Kohn, and Carmi Schooler. 1982. "Educational Self-Direction and the Cognitive Functioning of Students." *Social Forces* 63:923-44.

Mirowsky, John. 1987. "The Psycho-Economics of Feeling Underpaid: Distributive Justice and the Earnings of Husbands and Wives." *American Journal of Sociology* 92:1404-34.

Mortimer, Jeylan T., and Michael D. Finch. 1986a. "The Development of Self-Esteem in the Early Work Career." *Work and Occupations* 13:217-39.

____. 1986b. "The Effects of Part-Time Work on Adolescent Self-Concept and Achievement." Pp. 66-89 in *Becoming a Worker,* edited by Kathryn Borman and Jane Reisman. Ablex.

Mortimer, Jeylan T., Michael D. Finch, and Donald Kumka. 1982. "Persistence and Change in Development: The Multidimensional Self-Concept." Pp. 263-313 in *Life-Span Development and Behavior*. Vol. 4, edited by Paul D. Baltes and Orville G. Brim Jr. Academic Press.

Mortimer, Jeylan T., Michael D. Finch, and Geoffrey Maruyama. 1988. "Work Experience and Job Satisfaction: Variation by Age and Gender." Pp. 109-55 in *Work Experience and Psychological Development through the Life Span*, edited by Jeylan T. Mortimer and Kathryn M. Borman. Westview Publishing.

Mortimer, Jeylan T., Michael D. Finch, Michael J. Shanahan, and Seongryeol Ryu. 1992. "Work Experience, Mental Health, and Behavioral Adjustment in Adolescence." *Journal of Research on Adolescence* 2:25-57.

Mortimer, Jeylan T., and Jon Lorence. 1979a. "Work Experience and Occupational Value Socialization: A Longitudinal Study." *American Journal of Sociology* 84:1361-85.

____. 1979b. "Occupational Experience and the Self-Concept: A Longitudinal Study." *Social Psychology Quarterly* 42:307-23.

____. 1981. "Self-Concept Stability and Change from Late Adolescence to Early Adulthood." Pp. 5-42 in *Research in Community and Mental Health*. Vol. 2, edited by Roberta G. Simmons. JAI Press.

____. 1995. "Social Psychology of Work." Pp. 497-523 in *Sociological Perspectives on Social Psychology*, edited by Karen S. Cook, Gary Alan Fine, and James S. House. Allyn and Bacon.

Mortimer, Jeylan T., Jon Lorence, and Donald Kumka. 1986. *Work, Family, and Personality: Transition to Adulthood*. Ablex.

Mortimer, Jeylan T., and Michael J. Shanahan. 1994. "Adolescent Work Experience and Family Relationships." *Work and Occupations* 21:369-84.

Mortimer, Jeylan T., Michael J. Shanahan, and Seongryeol Ryu. 1993. "The Effect of Adolescent Employment on School Related Orientation and Behavior." Pp. 304-26 in *Adolescence in Context: The Interplay of Family, School, Peers and Work in Adjustment*, edited by Rainer K. Silbereisen and Eberhard Todt. Springer-Verlag.

Mortimer, Jeylan T., and Roberta G. Simmons. 1978. "Adult Socialization." *Annual Review of Sociology* 4:421-54.

Moss, Howard A., and Elizabeth J. Sussman. 1980. "Longitudinal Study of Personality Development." Pp. 530-95 in *Constancy and Change in Human Development*, edited by Orville G. Brim Jr. and Jerome Kagan. Harvard University Press.

O'Brien, Gordon E. 1986. *Psychology of Work and Unemployment*. John Wiley & Sons.

Olsen, Nancy J. 1974. "Family Structure and Socialization Patterns in Taiwan." *American Journal of Sociology* 79:1395-1417.

O'Malley, Patrick M., and Jerald G. Bachman. 1979. "Self-Esteem and Education: Sex and Cohort Comparisons among High School Seniors." *Journal of Personality and Social Psychology* 7:1153-59.

Owens, Timothy J. 1992a. "The Effect of Post-High School Social Context on Self-Esteem." *Sociological Quarterly* 33:553-77.

____. 1992b. "Where Do We Go from Here? Post-High School Choices of American Men." *Youth and Society* 23:452-77.

____. 1993. "Accentuate the Positive — and the Negative: Rethinking the Use of Self-Esteem, Self-Deprecation, and Self-Confidence." *Social Psychology Quarterly* 56:288-99.

____. 1994. "Two Dimensions of Self-Esteem: Reciprocal Effects of Positive Self-Worth and Self-Deprecation on Adolescent Problems." *American Sociological Review* 59:391-407.

Pfeffer, Jeffrey. 1981. "Management as Symbolic Action: The Creation and Maintenance of Organizational Paradigms." Pp. 1-52 in *Research in Organizational Behavior*. Vol. 3, edited by L.L. Cummings and Barry M. Straw. JAI Press.

Portes, Alejandro, and Kenneth L. Wilson. 1976. "Black-White Differences in Educational Attainment." *American Sociological Review* 41:414-31.

Purkey, William Watson. 1970. *Self-Concept and School Achievement*. Prentice-Hall.

Raven, John C. 1951. *Guide to Using Progressive Matrices (1947) Sets A, Ab, B.* Lewis.

Rehberg, Richard A., and Evelyn R. Rosenthal. 1978. *Class and Merit in the American High School.* Longman.

Rodin, Judith, Christine Timko, and Susan Harris. 1985. "The Construct of Control: Biological and Psychological Correlates." *Annual Review of Gerontology and Geriatrics* 5:3-55.

Rogers, Carl M., Monte D. Smith, and J. Michael Coleman. 1978. "Social Comparison in the Classroom: The Relationship between Academic Achievement and Self-Concept." *Journal of Educational Psychology* 70:50-77.

Rollins, Boyd C., and Darwin L. Thomas. 1979. "Parental Support, Power, and Control Techniques in the Socialization of Children." Pp. 317-64 in *Contemporary Theories about the Family.* Vol. 1, *Research-Based Theories,* edited by Wesley R. Burr, Reuben Hill, F. Ivan Nye, and Ira L. Reiss. Free Press.

Rosenberg, Morris. 1979. *Conceiving the Self.* Basic Books.

____. 1981. "The Self-Concept: Social Product and Social Force." Pp. 593-624 in *Social Psychology: Sociological Perspectives,* edited by Morris Rosenberg and Ralph H. Turner. Basic Books.

Rosenberg, Morris, and Leonard I. Pearlin. 1978. "Social Class and Self-Esteem among Children and Adults." *American Journal of Sociology* 84:53-77.

Rosenberg, Morris, Carmi Schooler, and Carrie Schoenbach. 1989. "Self-Esteem and Adolescent Problems: Modeling Reciprocal Effects." *American Sociological Review* 54:1004-18.

Schooler, Carmi, Joanne Miller, Karen A. Miller, and Carol N. Richtand. 1984. "Work for the Household: Its Nature and Consequences for Husbands and Wives." *American Journal of Sociology* 90:97-124.

Sears, David O. 1981. "Life-Stage Effects on Attitude Change, Especially among the Elderly." Pp. 183-204 in *Aging: Social Change,* edited by Sara B. Kiesler, James N. Morgan, and Valerie Kincade Oppenheimer. Academic Press.

Seashore, Stanley E., and Thomas D. Taber. 1975. "Job Satisfaction Indicators and Their Correlates." *American Behavioral Scientist* 18:333-68.

Sewell, William H., and Robert M. Hauser. 1975. *Education, Occupation, and Earnings: Achievement in the Early Career.* Academic Press.

Shanahan, Michael J., Michael D. Finch, Jeylan T. Mortimer, and Seongryeol Ryu. 1991. "Adolescent Work Experience and Depressive Affect." *Social Psychology Quarterly* 54:299-317.

Skinner, Ellen A. 1995. *Perceived Control, Motivation, and Coping*. Sage.

Slomczynski, Kazimierz, Joanne Miller, and Melvin L. Kohn. 1981. "Stratification, Work, and Values: A Polish-United States Comparison." *American Sociological Review* 46:720-44.

Spenner, Kenneth I., and Luther B. Otto. 1985. "Work and Self-Concept: Selection and Socialization in the Early Career." Pp. xx-xx in *Research in Sociology of Education and Socialization.* Vol. 5, edited by Alan C. Kerckhoff. JAI Press.

Steinberg, Laurence D., Suzanne Fegley, and Sanford M. Dornbusch. 1993. "Negative Impact of Part-time Work on Adolescent Adjustment: Evidence from a Longitudinal Study." *Developmental Psychology* 29:171-80.

Steinberg, Laurence, Nina S. Mounts, and Susie D. Lamborn. 1991. "Authoritative Parenting and Adolescent Adjustment across Varied Ecological Niches." *Journal of Research on Adolescence* 1:19-36.

Strickland, Bonnie R. 1989. "Internal-External Control Expectancies: From Contingency to Creativity." *American Psychologist* 44:1-12.

Super, Donald E. 1957. "The Multifactor Tests: Summing Up." *Personnel and Guidance Journal* 36:17-20.

Swann, William B. Jr., Alan Stein-Seroussi, and R. Brian Giesler. 1992. "Why People Self-Verify." *Journal of Personality and Social Psychology* 62:392-401.

Van Maanen, John, and Edgar Schein. 1979. "Toward a Theory of Organizational Socialization." *Research in Organizational Behavior* 1:209-64.

Weiner, Bernard. 1979. "A Theory of Motivation for Some Classroom Experiences." *Journal of Educational Psychology* 71:3-25.

Woelfel, Joseph, and Archibald O. Haller. 1971. "Significant Others, the Self-Reflexive Act and the Attitude Formation Process." *American Sociological Review* 36:74-87.

Wylie, Ruth. 1979. *The Self-Concept*. Rev. ed. Vol. 2. University of Nebraska Press.

Part-Time Work and Occupational Value Formation in Adolescence

JEYLAN T. MORTIMER, *University of Minnesota*
ELLEN EFRON PIMENTEL, *University of Minnesota*
SEONGRYEOL RYU, *University of Minnesota*
KATHERINE NASH, *University of Minnesota*
CHAIMUN LEE, *University of Minnesota*

Abstract

This study examines the interrelations of work experiences and occupational values in a panel of 1,000 ninth-graders followed over a four-year period. Neither employment itself nor hours of work had substantial effects on occupational value formation. However, the opportunity to learn useful skills at work was a consistent positive influence on increasingly stable intrinsic and extrinsic value dimensions. We conclude that the central finding of research on work and adult psychological functioning — that the conditions of work are what matter — can be generalized to adolescents.

Given that work values are important determinants of vocational choice (Davis 1965; Mortimer 1974) and actual occupational destinations (Mortimer, Lorence & Kumka 1986), it is important to understand their origins. This article addresses a potentially important, but thus far neglected, influence — adolescent work experience. American teenagers increasingly hold part-time jobs while attending school (Manning 1990). Because adolescence is a crucially important period for the formation of vocational interests and identity (Erikson 1968), it is reasonable to suppose that work experience would have a significant formative influence on adolescents' thinking about the potential rewards to be obtained from work.

Steinberg et al. (1981) posit three possible consequences of adolescent employment for vocational socialization. First, part-time jobs may help adolescents to acquire attitudes, values, habits, and knowledge that make them

** This article was presented at the 1992 meeting of the American Sociological Association. The study was supported by the National Center for Research on Vocational Education and by the National Institute of Mental Health (MH42843, "Work Experience and Adolescent Well-Being"). The authors would like to thank Michael D. Finch for his valuable consultation and Carol Krauze for bibliographic assistance.* *Direct all correspondence to Jeylan T. Mortimer, 1014 Social Sciences Building, University of Minnesota, 267 19th Avenue South, Minneapolis, Minnesota 55455.*

better adult workers. In contrast to this "competence model," the "cynicism model" proposes that since jobs available to teenagers are often menial, employed youth become cynical about work and focus on its extrinsic rewards. The third plausible hypothesis is that early work experience has no meaningful influence because teenagers view their jobs as temporary and do not expect to engage in similar work after completing school. Steinberg et al.'s (1981) finding, that part-time employment had little association with occupational values among 531 tenth- and eleventh-grade California high school students, supports the no-influence hypothesis.

Research on adults shows that the *conditions of work* are what matter for psychological development (Kohn & Schooler 1983; Mortimer, Lorence & Kumka 1986). If the findings of such research are applicable to younger workers, we should examine the quality of adolescents' work as well. However, past research on the psychosocial impacts of adolescent work experience has given far more attention to employment status (employed versus not) and to the intensity (hours) of work than to the quality of work. (For exceptions, see Stern et al. 1990; Wijting et al. 1977).

Given the paucity of research on the quality of adolescent work as a factor in occupational reward value formation, we look to studies of young adults for guidance. In Mortimer and Lorence's (1979) research on a panel of male college graduates, high levels of economic reward enhanced the valuation of extrinsic rewards. In this study, and in Lindsay and Knox's (1984) replication with a national sample of high school seniors, high monetary remuneration fostered declines in intrinsic values, while autonomous and challenging work strengthened intrinsic values over time.

However, Mortimer, Dennehy, and Lee (1991) find that self-direction at work has no significant relation to adolescent occupational values. Then what qualities of youth jobs might make a difference? In anticipating their adult "possible selves" (Markus, Cross & Wurf 1990), adolescents may be particularly responsive to occupational experiences that augur successful role adaptation, such as the acquisition of general proficiencies or skills. Being able to help others may also foster a satisfying sense of movement toward adult status, increasing the adolescent's sense of usefulness and competence.

In contrast, it is reasonable to suppose that work offering high levels of extrinsic reward (e.g., good pay) would foster extrinsic work values. Stressors in the workplace (House 1980; Kahn 1981; Kohn & Schooler 1983) could also foster an extrinsic orientation toward work. That is, if a young person's work experience is punishing, work may come to be seen not as providing intrinsic gratifications of value in itself, but only as a source of income and instrumental benefits.

Among adults, prior occupational values have been found to influence the selection of work, including the extent of autonomy, income, and social content (Lindsay & Knox 1984; Mortimer & Lorence 1979). However, there is reason to expect divergence in such person-job dynamics in this younger group, for whom occupational values may have less salience and motivational force. Moreover, because adolescents have a narrow array of work opportunities, concentrated in the retail and service sectors and, in the earlier years of high

school, in informal work, they may not be able to select jobs that are congruent with their values.

In addition to examining the relations between work and occupational values, we assess differences related to social background and gender. Adults in more advantaged social positions (i.e., white and of high educational and occupational attainment) have been found to be more intrinsically oriented; the less advantaged emphasize extrinsic values (Quinn, Staines & McCullough 1974; Shapiro 1977). Similar patterns have been observed among adolescents (Krau 1987; Lindsay & Knox 1984; Mannheim 1988). In fact, the effects of socioeconomic standing and race on occupational values may grow stronger over time, as adolescents make choices regarding academic curriculum that influence future occupational options. Parents and adolescent children may become increasingly involved in discussions about work, at which time parents transfer their work values to their children, especially when they have close, communicative relationships (Mortimer & Kumka 1982).

While the literature yields mixed findings on gender differences in occupational values, some studies show that men are more concerned with autonomy and the extrinsic benefits of jobs, whereas women emphasize more intrinsic concerns, such as having interesting work, opportunities to help others, and good interpersonal relationships (Mortimer & Lorence 1995). But because gender role norms are changing along with women's labor force behavior, contemporary boys' and girls' orientations to work may be more similar than in previous cohorts (Hanson 1988; Lueptow 1980). We also include grade-point average in the analysis.

Data Source

The data were obtained from a four-year longitudinal study (1988-91) of the effects of adolescent work experience on mental health and vocational development. The panel was chosen randomly from a list of ninth graders enrolled in the St. Paul (Minnesota) School District. A probit analysis of the census tract characteristics of general sample participants (64% of eligible cases) and those who chose not to participate indicated that the panel is representative of students attending the St. Paul schools (Finch et al. 1991; Mortimer et al. 1992).

Self-administered questionnaires were distributed in school classrooms each year (grades 9-12); students not present for either of two scheduled administrations (and those not attending school — 10% in wave 4) were mailed questionnaires, using procedures recommended by Dillman (1983). Of the initial 1,000 participants, 93% were retained over the four-year period.

MEASURES

In each wave the adolescents were asked, "In the future, how important do you think each of the following will be to you when you are looking for a FULL-TIME job?" Students rated specific features of work on a four-point scale, ranging from not at all important to extremely important. Four items reference extrinsic values: income, advancement opportunity, security, and prestige. Intrinsic values refer to the use of skills and abilities; learning opportunities;

autonomy and responsibility; and the chance to work with, and be helpful to, other people (see Appendix A). Confirmatory factor analyses showed that a two-factor model, with intrinsic and extrinsic dimensions, provided a good fit to the data in separate analyses for boys and girls in each grade (fit ratios, i.e., chi square divided by the degrees of freedom, ranged from 2.69 to 4.43; GFIs ranged from .936 to .966). Multiple group comparisons showed that the measurement structure for boys and girls was substantially the same.

The opportunity to obtain skills on the job and job stressors are latent variables defined by multiple indicators; income evaluation[1] and work with people are based on single indicators (see Appendix A). The analyses also include gender (coded 1 if male, 0 if female), race (coded 1 if white, 0 if other), grade-point average, and socioeconomic status (a composite of standardized measures of parental education and family income).

Findings

EMPLOYMENT STATUS, INTENSITY, AND OCCUPATIONAL VALUES

Both boys and girls have substantial work experience during high school. Forty percent of boys were employed in grade 9 in comparison to 63% of girls. The percentage of employed boys increased to 58% by wave 4; girls' employment rate increased to 70%. Ninth-grade boys and girls who had jobs both worked, on the average, about 11 hours per week. By the fourth wave, work intensity also increased for both — to 22 and 20 hours respectively.

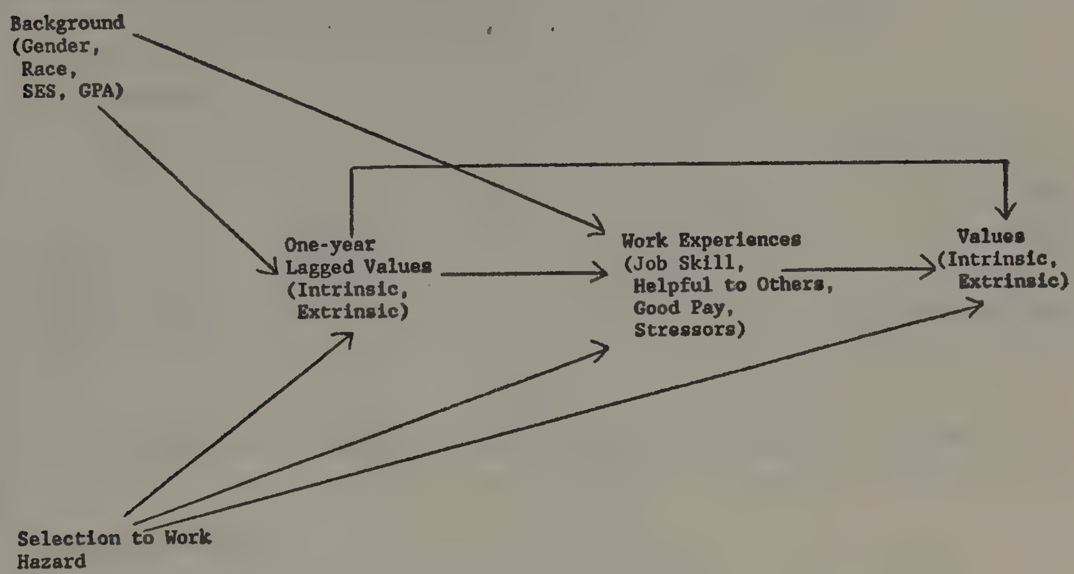
Using indices based on summations of the unweighted value items, we found no differences in intrinsic and extrinsic values, in any year, between boys who were employed and those who were not. Employed girls in grade 9 had stronger ($p < .01$) intrinsic values than their unemployed counterparts; and in wave 4, they had both higher mean intrinsic ($p < .01$) and extrinsic values ($p < .05$). There were no significant differences, by employment status, in girls' values in waves 2 and 3. Finally, neither intrinsic nor extrinsic occupational values were significantly correlated with hours worked in any wave for employed boys or girls.

Thus, there is no consistent evidence that the mere fact of employment during high school fosters distinct intrinsic or extrinsic occupational values. What evidence there is suggests that working may be associated with a higher rather than a lower intrinsic value orientation for girls. That is, in two of the four waves of data collection, girls who were employed placed greater emphasis than girls who were not employed on the inherent features of work, such as autonomy and having opportunities to use skills, learn new things, and work with people. There is no evidence that work intensity, the number of hours worked per week, influences occupational value formation.

MULTIVARIATE ANALYTIC STRATEGY

The hypothetical model shown in Figure 1 parallels those estimated by Lindsay and Knox (1984) and Mortimer and Lorence (1979). We hypothesize that the work conditions contemporaneously influence occupational values (Kohn &

FIGURE 1: Hypothetical Model



Schooler 1983; Mortimer, Lorence & Kumka 1986), while one-year lagged intrinsic and extrinsic values influence subsequent work experiences. Finally, the stability of intrinsic and extrinsic value constructs over time is tested.

There are three background variables whose influences on the value and work constructs are explored. We estimate the effects of gender,[2] race, socioeconomic status, and grade-point average on the one-year lagged values and work experience. A selection-to-work hazard rate influences all endogenous variables. If the determinants of employment were systematically related to occupational values or to work features, estimates could be biased if these bases of selection were not taken into account (Berk 1983). Therefore, three probit analyses were used to predict employment each year, with gender, age, race, family composition, socioeconomic status, and an indicator of the propensity to work[3] as predictors. These variables are temporally prior to both work experiences and values at each wave. Separate hazard rates, reflecting these predictors, are used to control selection to work at each wave (Heckman 1976, 1979).

This model is estimated three times, considering first the process of value formation between waves 1 and 2, second between waves 2 and 3, and finally between waves 3 and 4. In this way we explore developmental trends in occupational values and work experiences as students move through high school. Because different individuals work at each time, we do not attempt to model value development with the same cases over the entire period of study.

Latent variable measurement parameters and causal parameters are estimated simultaneously. While there is some variation in measurement parameters across equations, differences are slight. To indicate the relative importance of items for each construct, standardized lambda coefficients for second-wave latent variables are given in Appendix A. To simplify the models,

TABLE 1: Unstandardized Parameters for Models of Work Value Development

A. Effects of Exogenous Variables on Lagged Values[a]

	Intrinsic			Extrinsic		
	1	2	3	1	2	3
Gender	-.203***	-.092*	-.142*	—	—	—
Race	—	—	—	—	—	—
SES	ns	.034*	ns	ns	ns	-.027*
GPA	ns	ns	ns	ns	ns	ns
Selection hazard	ns	ns	ns	ns	ns	ns

B. Effects of Exogenous Variables and Lagged Values on Work Experiences

	Job Skill			Helpful to Others		
	2	3	4	2	3	4
Gender	-.320***	ns	ns	-.324**	ns	-.314**
Race	ns	-.194*	ns	—	—	—
SES	ns	ns	ns	.067*	ns	ns
GPA	-.098*	ns	ns	ns	ns	ns
Selection hazard	ns	ns	ns	ns	ns	ns
Lagged intrinsic	ns	.316***	.375***	ns	.301**	ns
Lagged extrinsic	—	—	—	—	—	—

	Good Pay			Stressors		
	2	3	4	2	3	4
Gender	—	—	—	.267*	.303**	.691***
Race	—	—	—	—	—	—
SES	—	—	—	ns	ns	-.077**
GPA	.120†	ns	ns	-.217***	ns	-.136*
Selection hazard	ns	ns	ns	.785*	ns	-.810**
Lagged intrinsic	—	—	—	ns	ns	.210**
Lagged extrinsic	—	—	—	—	—	—

all statistically insignificant ($p > .10$) paths were set to zero and the models reestimated. These reduced models are presented.

MULTIVARIATE FINDINGS

Unstandardized parameter estimates (obtained from the LISREL program, version 7.16-PC), considering occupational value formation across the three periods of study (between waves 1 and 2, 2 and 3, and 3 and 4), are shown in

TABLE 1: Unstandardized Parameters for Models of Work Value Development (Continued)

C. Effects of Selection to Work Hazard, Lagged Values, and Work Experiences on Values

	Intrinsic Values			Extrinsic Values		
	2	3	4	2	3	4
Selection hazard	ns	ns	-.324**	ns	ns	-.163*
Lagged intrinsic	.490***	.713***	.642***	—	—	—
Lagged extrinsic	—	—	—	.520***	.491***	.694***
Job skill	.120*	.218***	.100*	.074*	.070**	.078**
Helpful to others	.081**	ns	ns	—	—	—
Good pay	—	—	—	—	—	—
Stressors	—	—	—	.046*	ns	ns

D. Correlated Residuals[b]

	2	3	4
Job skill and helpful to others	.205*** (.369)	.174*** (.310)	.146*** (.260)
Lagged intrinsic and lagged extrinsic values	.092*** (.737)	.106*** (.629)	.079*** (.511)
Intrinsic and extrinsic values	.062*** (.491)	.072*** (.387)	.058*** (.356)

[a] Dashes indicate paths that were dropped in the final model.
[b] Standardized parameters are in parentheses.

† $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

Table 1 to facilitate comparisons across years. Each model represents a good fit to the data, with chi square values divided by the degrees of freedom of 1.82, 1.99, and 1.89 for each period, respectively.

WORK VALUES AND EXPERIENCES

Most striking is the consistent positive effect of the construct we call job skills, signifying opportunity to learn skills on the job that are useful in the future, on the intrinsic value dimension (see Table 1, panel C). When adolescents perceive their jobs as allowing them to move forward on their occupational trajectories — enabling them to learn how to accept responsibility, get along with people, be on time, manage money, and follow directions — they come to view the

intrinsic benefits of work as more important. Of three tests of this effect, all coefficients are positive and statistically significant. However, the effect of job skills on intrinsic values decreases after wave 3 ($z = 1.94$, $p < .10$).

Given that we control the substantial stability of values, the findings provide convincing evidence that the opportunity to acquire skills on the job strengthens adolescents' intrinsic orientation toward work. Moreover, the effect of job skill acquisition extends beyond intrinsic values to influence the extrinsic construct as well. Significant positive coefficients, of similar magnitude, are evident in all periods. Lindsay and Knox (1984) likewise report what they call a crossover effect of intrinsic work features on extrinsic values.

With respect to the other work conditions, having opportunity to be helpful to others positively influences intrinsic values in the first wave; it has no significant effects subsequently. Also, only in the first wave does work stress have a significant positive effect on extrinsic values. Thus, there is a tendency, in just one period, for younger workers who experience more stress on the job to come to view work as a source of extrinsic occupational rewards. Pay has no significant impact on extrinsic values.

It could be argued that even though we have controlled prior values in estimating the effects of work, there are contemporaneous processes of reverse causation that are not taken into account. For example, persons with higher levels of intrinsic values could perceive their jobs as offering more opportunities to develop their skills. They might also mold their jobs so as to realize their values.

Because of a lack of adequate identification, we were unable to assess all pertinent reciprocal effects. However, we did estimate several models specifying reciprocal causal pathways. They provided no evidence for a pattern of reverse causation. Paths from wave 2 intrinsic values to wave 2 job skills and helpfulness, and likewise from extrinsic values to stressors, were not statistically significant. The path from wave 4 extrinsic values to job skills was also insignificant. We were unable to estimate the effects of instrinsic values in waves 3 and 4 on contemporaneous job skills because of poor model identification.

Consistent with our speculation that adolescents are unlikely to be able to choose their work on the basis of their values, given the relatively small selection of jobs from which they have to choose, wave 1 occupational value orientations have no significant effects on wave 2 work experiences (see Table 1, panel B). However, this situation changes as adolescents mature.

Wave 2 intrinsic values have a significant, positive effect on the report the following year that one's job enables the development of useful skills; this pattern is replicated one year later. Intrinsic values measured in wave 2 also significantly predict wave 3 opportunities to be helpful to others, although this effect is not repeated in the final period. Statistical comparison of coefficients representing the effects of lagged intrinsic values on subsequent opportunities to develop job skills shows a significant change between waves 1 and 2. However, the coefficients do not significantly increase after the second period.[4]

It is somewhat anomalous that intrinsic values in wave 3 also have a positive effect on wave 4 work stressors. Some characteristics that make work intrinsically rewarding (e.g., responsibility and self-direction) may also be

stressful to occupational newcomers (see Shanahan et al. 1991). The more intrinsically motivated adolescents may steer themselves toward work that is challenging but also more stressful. Extrinsic values, however, were found to have no bearing on subsequent work experiences.

Also noteworthy is the increasing net stability of occupational values over the high school years. The unstandardized stability paths for extrinsic values rise from .520 ($p < .001$) between waves 1 and 2 to .694 ($p < .001$) between waves 3 and 4 (see Table 1, panel C), with stability in the last period marginally greater than in the first period ($z = 1.88$, $p < .10$). Intrinsic value stability significantly increases between the first and second periods only ($z = 2.24$, $p < .05$).

THE EFFECTS OF BACKGROUND VARIABLES ON WORK VALUES

Whereas boys are less intrinsically oriented than girls in all periods, there are no gender differences in extrinsic values (see Table 1, panel A). There are no significant differences between the racial categories in values.

There is a tendency for socioeconomic background to be more positively related to adolescents' intrinsic values in the latter than in the earlier periods of study. It manifests no significant influence in wave 1 but has significant positive effects on intrinsic values in wave 2 and also in 4 ($\beta = .046$, $\beta = .141$, $p < .001$, not shown), controlling race, gender, GPA, and the selection-to-work hazard. (Although the effect of SES on intrinsic values in waves 2 and 4 is statistically similar, both are significantly different from wave 1, with $z = 2.21$ and 2.82 respectively). With the same controls, social class background bears a significant, negative relationship to third-wave extrinsic values, but insignificant relationships in waves 1 and 2 (and wave 4, not shown).[5] GPA has no consistent effect on occupational values.

Positive residual correlations between the intrinsic and extrinsic value factors in each period indicate that adolescents who have higher intrinsic orientations are also the more extrinsically oriented (see Table 1, panel D). Jobs offering opportunities for skill development and to help others also appear to go together.

Summary and Conclusions

Little attention has been given in prior research to the effect of adolescent conditions of work — as opposed to the fact of work or its intensity — on the formation of occupational values. We do not find that work status itself, or the intensity of adolescent work, have consistent effects on occupational value formation. But there is substantial grounding for the conclusion that the quality of adolescents' jobs, particularly the challenges and demands that enable the development of skills, influences intrinsic and extrinsic work values. We conclude that the central finding of research on the effects of work on adult psychological functioning — that the conditions of work are what matter — can be generalized to adolescents.

In terms of the three models noted earlier — competence, cynicism, and no influence of work on values — our analysis supports a competence model.

Despite being employed largely in jobs of low status and complexity, adolescents still gain a sense of appropriate developmental progress from learning new skills, sufficient to develop stronger intrinsic value orientations. Opportunities to develop job skills fostered stronger extrinsic values as well.

It is only among tenth-graders that the chance to be helpful to others reinforces intrinsic work orientations, and that work stressors show a tendency to foster extrinsic values. Opportunities to be helpful may be considered markers of the transition to adult roles, or "possible selves," which foster positive attitudes toward work and intrinsic occupational values. Stressors may have less impact on occupational value formation as the adolescent learns to cope with them, through experience in jobs that hone their work skills. Job skills continue to strengthen both intrinsic and extrinsic values across time. It may be that both intrinsic and extrinsic benefits are seen as increasingly within reach given successful learning and adaptation to the demands of adolescent work.

However, there appears to be a developmental pattern in the effect of job skills on intrinsic value formation. While the effect is significant in all three periods, it significantly decreases after wave 3. Similarly, there are diminishing effects of helpfulness and stress on values after wave 2. As values become a more enduring characteristic of the individual, work characteristics may play a decreasing role in their articulation, though the process is by no means finished in high school. Still, the job skills index has an unchanging effect on extrinsic values over the period studied.

There are indications of an emergent relationship between social class background and values. While our analyses partially support a link between more advantaged socioeconomic status, stronger intrinsic values, and weaker extrinsic values, there are no value differences by race. Girls are found to be more intrinsically oriented. The absence of a gender difference in extrinsic values may result from social change in gender norms. Alternatively, gender differences in the development of extrinsic values may appear only later in life.

Our findings thus suggest developmental dynamics such that the relations among values, work experiences, and individual differences become increasingly similar to those manifest among adults as adolescents move through high school. As hypothesized, values that the youngest adolescents hold have no significant effect on subsequent work characteristics. In the earliest transition to work, job conditions are not significantly affected by adolescents' values about work. However, as they move through high school, young people's intrinsic values increasingly shape the kind of work they do, at the same time as they are presented with greater job choices.

Adolescents' occupational values also become significantly more stable across time. Previous research (Lorence & Mortimer 1985; Mortimer, Finch & Maruyama 1988) has documented the greater stability of other work orientations (i.e., job involvement and job satisfaction) in older workers, as compared to those who are younger (age 16-25). This pattern was attributed to processes of "aging stability": because of distinct environmental pressures and developmental processes, attitudes stabilize with age. (See Alwin, Cohen & Newcomb 1991 for similar findings with respect to political attitudes.) However, this research demonstrates that a pattern of "aging stability" in occupational values extends to the youngest workers.

APPENDIX A: Measures[a]

Extrinsic Value Construct

- Good pay (.281)
- Good chances of getting ahead (.386)
- A steady job, with little chance of being laid off (.446)
- A job that people regard highly (.454)

Intrinsic Value Construct

- A job that uses my skills and abilities (.449)
- A chance to learn a lot of new things at work (.463)
- A chance to make my own decisions at work (.500)
- A job where I have a lot of responsibility (.512)
- A chance to be helpful to others or useful to society[b] (.416)
- A chance to work with people rather than things[b] (.485)

WORK EXPERIENCES

Job Skills

- How much has your job helped you to develop the following abilities? (1 = not at all, to 4 = a great deal)
- To follow directions (.595)
- To get along with others (.638)
- To be on time (.736)
- To take responsibility for your work (.705)
- To manage your money (.456)
- Do you think that the things you are learning in your job will be useful to you in your later life? (1 = not at all useful, to 4 = extremely useful) (.421)

Helpful to Others

- My job gives me a chance to be helpful to others (1 = not at all true, to 4 = very true).

Pay Evaluation

- Would you consider your pay "good pay" for the work you do? (1 = no, the pay is not good, to 4 = yes, definitely)

Stressors

- Response categories for the first five items: 1 = not at all true, to 5 = very true
- I have too much work to do everything well (.821)
- My job requires that I work very hard (.883)
- I feel drained of my energy when I get off work (.574)
- How often is there time pressure on your job? (1 = never, to 5 = almost always) (.798)
- How often are you exposed to excessive heat, cold, or noise at work? (1 = never, to 5 = almost always) (.727)

[a] Standardized lambdas, estimated simultaneously with the causal parameters, are given in parentheses for each multiple-indicator latent variable construct, measured in grade ten.

[b] The residuals of these items, indicating people-oriented concerns, were significantly related; the unstandardized covariance is .158 ($p < .001$).

By continuing to follow this panel as it makes the transition to adulthood, we will be able to examine the dynamics of change in values following high school in response to work and other experiences, and the effects of occupational value orientations on actual attainments in early adulthood.

Notes

1. There is evidence that evaluation of income has a stronger impact on work orientations than actual income (Gruenberg 1980; Janson & Martin 1982; Jencks, Perman & Rainwater 1988; Mortimer et al. 1988). However, Lindsay and Knox (1984) report a negative effect of income on intrinsic values, and Mortimer and Lorence (1979) find that high income fosters extrinsic values. Of 16 correlations of hourly pay and values (2 values, 2 genders, and 4 waves) only one was statistically significant. In wave 4, girls with higher wages expressed stronger extrinsic values ($r = .12$, $p < .05$). However, this relationship became statistically insignificant with the lagged variable controlled.

2. Because prior analysis showed that the process of work value development was essentially the same for boys and girls (Mortimer, Dennehy & Lee 1991), the models were based on a variance-covariance matrix for all students.

3. The propensity to work was indicated by the age of starting one's first job. For students who never worked, the current age was assigned.

4. Coefficients were compared statistically using a z test, where $z = (B1 - B2)/(SE^2_{B1} + SE^2_{B2})^{1/2}$. Z scores comparing the effects of lagged intrinsic values on job skills are 2.06 (waves 2 and 3), 2.77 (waves 2 and 4), and .58 (waves 3 and 4).

5. Given this pattern, it could be argued that socioeconomic status should be controlled when assessing the effect of work on values. Therefore, we freed the parameters from socioeconomic status to the later value constructs in each model. Socioeconomic status did have significant effects on intrinsic values in waves 2 and 4. However, because none of the effects of the work experiences on intrinsic values was altered, we retained the original, simpler model specification.

References

Alwin, Duane F., Ronald L. Cohen, and Theodore M. Newcomb. 1991. *Political Attitudes over the Life Span*. University of Wisconsin Press.

Berk, Richard A. 1983. "An Introduction to Sample Selection Bias in Sociological Data." *American Sociological Review* 48:386-98.

Davis, James A. 1965. *Undergraduate Career Decisions: Correlates of Occupational Choice*. Aldine.

Dillman, Don A. 1983. "Mail and Other Self-Administered Questionnaires." Pp. 359-77 in *Handbook of Survey Research*, edited by Peter H. Rossi, James D. Wright, and Andy Anderson. Academic Press.

Erikson, Erik H. 1968. *Identity, Youth, and Crisis*. W.W. Norton.

Finch, Michael D., Michael J. Shanahan, Jeylan T. Mortimer, and Seongryeol Ryu. 1991. "Work Experience and Control Orientation in Adolescence." *American Sociological Review* 56:597-611.

Gruenberg, Barry. 1980. "The Happy Worker: An Analysis of Educational and Occupational Differences in Determinants of Job Satisfaction." *American Journal of Sociology* 86:247-71.

Hansen, Jo-Ida C. 1988. "Changing Interests of Women: Myth or Reality?" *Applied Psychology: An International Review* 37:133-50.

Heckman, James J. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models." *Annals of Economic and Social Measurement* 5:475-92.

____. 1979. "Sample Selection as a Specification Error." *Econometrica* 45:153-61.

House, James S. 1980. *Occupational Stress and the Mental and Physical Health of Factory Workers.* Survey Research Center, Institute for Social Research, University of Michigan.

Janson, Philip, and Jack K. Martin. 1982. "Job Satisfaction and Age: A Test of Two Views." *Social Forces* 60:1089-1102.

Jencks, Christopher, Lauri Perman, and Lee Rainwater. 1988. "What Is a Good Job? A New Measure of Labor Market Success." *American Journal of Sociology* 93:1322-57.

Kahn, Robert L. 1981. *Work and Health.* Wiley.

Kohn, Melvin L., and Carmi Schooler. 1983. *Work and Personality: An Inquiry into the Impact of Social Stratification.* Ablex.

Krau, Edgar. 1987. "The Crystallization of Work Values in Adolescence: A Sociocultural Approach." *Journal of Vocational Behavior* 30:103-23.

Lindsay, Paul, and William E. Knox. 1984. "Continuity and Change in Work Values among Adults: A Longitudinal Study." *American Journal of Sociology* 89:918-31.

Lorence, Jon, and Jeylan T. Mortimer. 1985. "Job Involvement through the Life Course: A Panel Study of Three Age Groups." *American Sociological Review* 50:618-38.

Lueptow, Lloyd B. 1980. "Social Change and Sex-Role Change in Adolescent Orientation toward Life, Work, and Achievement." *Social Psychology Quarterly* 43:48-59.

Mannheim, Bilha. 1988. "Social Background, Schooling, and Parental Job Attitudes as Related to Adolescents' Work Values." *Youth and Society* 19:269-93.

Manning, Wendy D. 1990. "Parenting Employed Teenagers." *Youth and Society* 22:184-200.

Markus, Hazel, Susan Cross, and Elissa Wurf. 1990. "The Role of the Self-System in Competence." Pp. 205-25 in *Competence Considered,* edited by Robert J. Sternberg and John Kolligian Jr. Yale University Press.

Mortimer, Jeylan T. 1974. "Patterns of Intergenerational Occupational Movements: A Smallest-Space Analysis." *American Journal of Sociology* 79:1278-99.

Mortimer, Jeylan T., Katherine Dennehy, and Chaimun Lee. 1991. "Work Experience and Other Influences on Adolescents' Occupational Values: Evidence from a Prospective Longitudinal Study." Paper presented at the American Vocational Association Convention, Los Angeles.

Mortimer, Jeylan T., Michael D. Finch, and Geoffrey Maruyama. 1988. "Work Experience and Job Satisfaction: Variation by Age and Gender." Pp. 109-55 in *Work Experience and Psychological Development through the Life Span,* edited by Jeylan T. Mortimer and Kathryn M. Borman. Westview Press.

Mortimer, Jeylan T., Michael D. Finch, Michael J. Shanahan, and Seongryeol Ryu. 1992. "Work Experience, Mental Health and Behavioral Adjustment in Adolescence." *Journal of Research on Adolescence* 2:25-57.

Mortimer, Jeylan T., and Donald Kumka. 1982. "A Further Examination of the 'Occupational Linkage Hypothesis.'" *Sociological Quarterly* 23:3-16.

Mortimer, Jeylan T., and Jon Lorence. 1979. "Work Experience and Occupational Value Socialization: A Longitudinal Study." *American Journal of Sociology* 84:1361-85.

____. 1995. "Social Psychology of Work." Pp. 497-523 in *Sociological Perspectives on Social Psychology,* edited by Karen S. Cook, James S. House, and Gary Alan Fine. Allyn and Bacon.

Mortimer, Jeylan T., Jon Lorence, and Donald S. Kumka. 1986. *Work, Family, and Personality: Transition to Adulthood.* Ablex.

Quinn, Robert P., Graham L. Staines, and Margaret R. McCullough. 1974. *Job Satisfaction: Is There a Trend?* Manpower Research Monograph no. 30. U.S. Department of Labor. Manpower Administration.

Shanahan, Michael J., Michael D. Finch, Jeylan T. Mortimer, and Seongryeol Ryu. 1991. "Adolescent Work Experience and Depressive Affect." *Social Psychology Quarterly* 54:299-317.

Shapiro, E. Gary. 1977. "Racial Differences in the Value of Job Rewards." *Social Forces* 56:21-30.

Steinberg, Laurence D., Ellen Greenberger, Alan Vaux, and Mary Ruggiero. 1981. "Early Work Experience: Effects on Adolescent Occupational Socialization." *Youth and Society* 12:403-22.

Stern, David, James R. Stone III, Charles Hopkins, and Martin McMillion. 1990. "Quality of Students' Work Experience and Orientation toward Work." *Youth and Society* 22:263-82.

Wijting, Jan P., Carole R. Arnold, and Kelly A. Conrad. 1977. "Relationships between Work Values, Socio-Educational and Work Experiences, and Vocational Aspirations of 6th, 9th, 10th, and 12th Graders." *Journal of Vocational Behavior* 11:51-65.

Demographic, Socioeconomic, and Behavioral Factors Affecting Ethnic Mortality by Cause*

RICHARD G. ROGERS, *University of Colorado*
ROBERT A. HUMMER, *Louisiana State University*
CHARLES B. NAM, *Florida State University*
KIMBERLEY PETERS, *University of Colorado*

Abstract

This article examines ethnic differences in total and cause-specific mortality. We employ the linked National Health Interview Survey-National Death Index (NHIS-NDI) to examine ethnic differences in mortality from a combination of demographic, socioeconomic, and health characteristic perspectives. We find that Asian American mortality is low in part because of healthy behaviors and socioeconomic advantages; that Caucasian American mortality is higher partly because of high prevalence and quantity of cigarette smoking; and that Mexican, Native, and African American mortality is higher partly from socioeconomic disadvantages. These results give us added insight into the demographic, social, and health mechanisms that lead us to persevere or to perish.

Most mortality research has focused on how demographic characteristics — age, sex, and ethnicity — affect total and cause-specific mortality. But to properly conceptualize the multiple and interacting proximate determinants of mortality by ethnicity, researchers must extend the traditional demographic framework of mortality analysis to a more complete, though more complex, multivariate model that also incorporates social characteristics — marital status, education, and income (see Mosley & Chen 1984; Rosenwaike 1988). Moreover, health behaviors have cultural dimensions and therefore can also affect ethnic differences in mortality. A new file, the linked National Health Interview Survey-National Death Index (NHIS-NDI), allows us to disentangle some of the effects of demographic, socioeconomic, and behavioral health characteristics on ethnic differences in total and cause-specific mortality.

* *This research was supported in part by National Institute on Aging Grant no. 1R01 AG 10113-01A2. The mortality data were collected by the National Center for Health Statistics. An earlier version of this article was presented at the 1995 annual meeting of the American Sociological Association, Washington, D.C. We thank Jay Teachman for his expert statistical advice and Diane Lye and two anonymous reviewers for their helpful comments and insights.* *Direct correspondence to Richard Rogers, Population Program, Campus Box 484, University of Colorado, Boulder, CO 80309.*

 Social Forces, June 1996, 74(4):1419-1438

Ethnicity (defined here broadly to include race and nationality characteristics — see Cooper 1991, 1994) is a fundamental demographic determinant of mortality. In the important ongoing debates over the measurement of ethnicity in relation to health outcomes (Cooper 1991; Crews & Bindon 1991; Williams, Lavizzo-Mourey & Warren 1994), we move away from the biological connotations of race and ethnicity to concentrate on the social significance of group membership (Dressler 1993). Thus, our research focuses on the broad range of factors that make ethnicity a powerful and socially meaningful concept in mortality research (Hummer 1996; Williams 1990; Williams, Lavizzo-Mourey & Warren 1994). By including in this study a wide range of covariates — demographic, socioeconomic, social relations, and health behaviors — of largely social character, we can capture some of the social significance of ethnicity and determine how variables within each set act as mechanisms by which ethnicity influences the mortality of individuals.

Demographic Perspective of Ethnic Differences in Mortality

Scholars have long acknowledged that ethnic differences in mortality are affected by age, sex, and medical cause of death. Life expectancies by ethnicity reveal marked differences. In 1980 Asians displayed the highest life expectancy, 82, which was even higher than that of Caucasian Americans, 74 (see Barringer, Gardner & Levin 1993). Mexican American life expectancy is similar to that of Caucasian Americans. Native American life expectancy, 71, is intermediate between that of Caucasian Americans and that of African Americans, at 68 (Rogers 1996; Snipp 1989). Although African American life expectancy lags behind that of other ethnic groups, it has increased the most dramatically from the turn of the century through the mid-1980s (Rogers 1996).

Variations in age-specific mortality rates suggest that some apparent differences may be partly a function of methodological problems. For example, Native Americans have lower life expectancy than Caucasian or Asian Americans, but lower age-specific mortality rates than Caucasians at ages 45 and above and lower than Asians at ages 85 and above. These low Native American age-adjusted mortality rates are most likely underestimates, reflecting a 45% increase in the reported Native American population between 1980 and 1990 and that in turn reflects the increased propensity for individuals to identify themselves as Native American (NCHS 1994b). The black/white gap in life expectancy is not consistent at all ages; instead, it is generally larger at lower ages and smaller at higher ages (Sorlie et al. 1992), and may even cross at the highest ages (Nam 1995).

Ethnic variations in mortality also result from differences in cause of death. Compared to Caucasians, Asians exhibit lower mortality for almost every cause of death, including death from cancer, human immunodeficiency virus infections (HIV), firearms, and motor vehicles (Fingerhut, Jones & Makuc 1994; NCHS 1994b). NCHS (1994b) has reported that compared to Caucasians of similar ages, homicide rates are lower for Asians at ages 15-24 and 25-44, although slightly higher for Asians at ages 45-64. Asians achieve their superior

life expectancies through reductions in both degenerative diseases and external causes.

Although overall Mexican American mortality is similar to that of Caucasian Americans, it results from different causes. Compared to Caucasian Americans, Mexican Americans are generally less likely to die from several of the major causes of death, including heart disease, cerebrovascular disease, cancer, pneumonia and influenza, suicide, and chronic obstructive pulmonary diseases and allied conditions (COPD); they are generally more likely to die from diabetes, homicide, cirrhosis of the liver, and HIV (NCHS 1994b).

Compared to Caucasians, Native Americans are less likely to die from cardiovascular diseases, cancer, and COPD, but more likely to die from diabetes, social pathologies, and infectious diseases. Historically, infectious diseases have been a major cause of death for Native Americans, whose tuberculosis rate is ten times as high as the rate for Caucasian Americans (U.S. Dept. of Health and Human Services [HHS] 1993). But now accidents and violence have replaced infectious diseases as major causes of death among Native Americans.

Compared to Caucasian Americans, African Americans die at higher rates from most causes of death, including circulatory diseases, cancers, respiratory diseases, infectious diseases, diabetes, homicide, and accidents (Kochanek, Maurer & Rosenberg 1994). Although most researchers have found that African Americans are less likely than Caucasian Americans to die from suicide, some studies indicate that figures for suicides are underestimated more often for African than Caucasian Americans (see Warshauer & Monk 1978). And compared to Caucasian Americans, African Americans are more than three times as likely to die from HIV and over six times as likely to die from homicide (Fingerhut, Jones & Makuc 1994; NCHS 1994a).

Social Characteristics Perspective

Ethnic patterns in mortality may reflect variability in a host of complex social and economic characteristics (Frisbie & Bean 1978). African Americans lag behind Caucasian Americans in life expectancy partly because African Americans are more likely to be in disadvantaged social categories. For instance, compared to Caucasian Americans, African Americans are less likely to be married, employed, or wealthy, characteristics that have been shown to be associated with lower mortality (Potter 1991; Schoenborn 1986). When only demographic variables are considered, compared to Caucasians, African Americans have higher total mortality and higher mortality from nearly all causes, save COPD and suicide. But controlling for social characteristics — marital status, family size, income, and household crowding — virtually eliminates the racial gap in total mortality and reduces the racial gap for most causes, including circulatory diseases, cancer, respiratory diseases, accidents, infectious diseases, diabetes, and homicide (Centerwall 1984; Rogers 1992; Sterling, Rosenbaum & Weinkam 1993). We can expect similar results for other socially disadvantaged ethnic groups.

For example, the "epidemiological paradox" (Markides & Coreil 1986) — that Mexican and Caucasian Americans have similar life expectancies, even

though a greater proportion of Mexican Americans live in poverty and lack health insurance (Sorlie et al. 1993) — hides social differences in the causes of death. Compared to that of Caucasian Americans, Mexican American mortality is higher for accidents and three times as high for homicide (personal communication, Jeff Maurer, NCHS). The high rates of homicide for young Mexican Americans is due in part to gang violence (see Loya & Mercy 1985). The high accident rate among Mexican Americans is due to a variety of factors but may include lower socioeconomic status, which may result from driving older, poorly maintained cars and living in areas that require more driving time. Low incomes may also increase the risk of nonautomotive accidents because low incomes are associated with deteriorating housing and overcrowding (Shai, Rosenwaike & Rogers 1991).

Compared to Caucasians, Asians enjoy lower overall mortality and lower mortality from most causes of death. And compared to other subpopulations, including Caucasians, Asians have some of the highest income and educational levels. For instance, median family income in 1992 was $43,400 for Asians, $39,300 for Caucasians, $24,900 for Hispanic Americans, and $21,800 for African Americans (U.S. Bureau of the Census 1994). These relations suggest that the low mortality experienced by Asians is due in part to their favorable socioeconomic status and the benefits that ensue from it.

Compared to Caucasians, Native Americans are more than three times as likely to die from cirrhosis of the liver and more than two and one-half times as likely to die from accidents. On a Hopi reservation in northeastern Arizona, rollover crashes, pedestrian fatalities, and falls — from pickup trucks, mesas, and pueblo roofs — contribute to high accident mortality (Simpson et al. 1983). Native Americans may suffer more fatal accidents because they live in more dangerous environments or because, in general, lower socioeconomic status increases a person's risk of accidents. The high ratio for cirrhosis of the liver suggests that alcohol abuse may contribute to the high rate of accidents. Indeed, Native Americans are more than six times as likely to die from alcoholism as Caucasian Americans (HHS 1993). Rates of suicide and homicide are also high among Native Americans. In fact, Native Americans are 1.6 times as likely as Caucasian Americans to die from firearms (Fingerhut, Jones & Makuc 1994). High rates of suicide have been documented among the Hopi, Apache, Papago, Shoshone-Bannock, Northern Cheyenne, and Blackfeet. And even these high rates may understate reality, because Native American cultural and religious taboos against suicide may lead to an underreporting of it (Simpson et al. 1983).

Behavioral Health Characteristics Perspective

Cigarette smoking is considered the single most important preventable cause of death (CDC 1989). Thus, ethnic smoking patterns may translate into ethnic differences in mortality. Compared to other ethnic groups, including Caucasians, Asian and Pacific Islander Americans report "the lowest use of cigarettes, alcohol, marijuana, and cocaine and other hard drugs" (Zane & Kim 1994:318).

Compared to Caucasians, Mexican Americans generally smoke less and therefore have lower mortality from such smoking-related diseases as heart

disease, cancer, and COPD (see Castro, Baezconde-Garbanati & Beltran 1985). Indeed, the age-adjusted rates of death (per 100,000 population) from lung cancer in the U.S. in 1980 were 48 for Caucasians and 59 for African Americans, but only 29 for the Mexican born (Rosenwaike 1987). Thus, the healthy behavior of Mexican Americans may compensate in part for their social disadvantages.

Circulatory disease, respiratory disease, and cancer, all smoking-related causes of death, contribute disproportionately to the black/white gap in mortality. In fact, cancer, the second leading cause of death, contributes almost one-quarter of the black male excess in mortality (Rogers 1992). Several factors help explain these differences. For example, African American males are more likely to smoke than Caucasian males (Jaynes & Williams 1989; Manton, Patrick & Johnson 1987), though among smokers, African American males smoke fewer cigarettes than do Caucasian Americans (Kabat, Morabia & Wynder 1991).

This overview of ethnic differences in mortality suggests two main hypotheses. First, the social characteristics hypothesis suggests that controlling for social and economic characteristics will reduce the overall and cause-specific mortality gaps between Caucasians and other ethnic groups. These relations will emerge in two ways. Because Caucasians enjoy lower mortality and greater rates of marriage, more education, and higher incomes than Mexican, Native, and African Americans, controlling for social characteristics will narrow the mortality gap between Caucasians and these ethnic groups. And because Asians exhibit lower mortality and higher levels of income and education than Caucasians, controlling for socioeconomic status will reduce the mortality gap between these two groups. Second, the health characteristics hypothesis, which focuses on cigarette smoking, asserts that controlling for smoking will also narrow the overall and cause-specific mortality gap between Caucasian and Asian Americans, will have little effect on the gap between Caucasian and African Americans, and will widen the gap between Caucasian and Mexican and Native Americans. These relations will emerge because Caucasian and African Americans have high rates of cigarette smoking and because Caucasians are more likely than other groups to smoke large quantities of cigarettes.

Data and Methods

LINKED NATIONAL HEALTH INTERVIEW SURVEYS AND NATIONAL DEATH INDEX

The National Center for Health Statistics has now linked the National Health Interview Survey (NHIS) and the National Death Index (NDI) for the years 1986 through 1991 to create the NHIS-NDI file, a unique and powerful data base for mortality research (NCHS 1994c). The NHIS-NDI matched data set, which is nationally representative, provides annual cross-sectional data for the non-institutionalized U.S. civilian population, includes detailed covariates, and allows an examination of those who are exposed to the risk of death over time. The linked data sets alleviate the problem of inconsistencies in ethnic identification between two disparate data sources.

The NHIS-NDI data set, assembled for those individuals 18 years old and older, provides detailed data for NHIS subjects who have survived subsequent years and also for those who have died. NCHS devised a probabilistic matching

scheme that assigns weights to each of twelve items — social security number; first name; last name; middle initial; race; sex; marital status; day, month, and year of birth; state of birth; and state of residence — to determine the quality of potential matches (Horm 1993; NCHS 1994c). Patterson and Bilgrad (1986) have demonstrated that the matching methodology is highly accurate.

NHIS data are cross-sectional and derived from a household multistage sampling of the civilian noninstitutional population of the U.S. (NCHS 1980). The NHIS includes two types of questions: the Basic Health and Demographic Questionnaire (BHD), composed of "core" questions that are asked annually, and the Current Health Topic Questionnaires (CHT), called supplements, which focus on sets of health issues that vary from year to year (see Chyba & Washington 1993). The core questions include basic demographic items such as age, ethnicity, sex, marital status, income, and education. A wide variety of supplemental questions are asked on current health topics, including cigarette smoking; some of the supplemental questions are asked of the entire sample and others are asked of subsamples.

Our analyses use the NHIS 1987 Cancer Epidemiology (*N* = 22,080), the 1987 Cancer Control (*N* = 22,043), the 1988 Occupational Health (*N* = 44,233), and the 1990 Health Promotion and Disease Prevention Supplements (*N* = 41,104) (NCHS 1992, 1993a, 1993b, 1994c). We selected these supplements because they contain questions on demographic, social, economic, and behavioral health, specifically smoking behavior. We matched the NHIS records to the corresponding deaths in the NHIS-NDI file (NCHS 1994c) and excluded less than 5% of the sample that was other than Asian, Native, African, Mexican, or Caucasian American to create a final sample of 398,204 person-year records (394,071 survival records and 4,133 death records). This sample is based on individuals who survived from 1987 through 1991 or who died sometime within the period. Because interviews for each sample were conducted throughout the year (in either 1987, 1988, or 1990), respondents were exposed to the risk of death for half of the interview year and for the entire year for each subsequent year. Thus, individuals who were interviewed in 1987 and who survived through 1991 contributed 4½ person-years, while those who died contributed person-years up to the time of death. Similarly, those who were interviewed in 1988 or 1990 and who survived through 1991 contributed 3½ and 1½ person-years, respectively.

ETHNIC IDENTIFICATION

Historically, it has been difficult to calculate national mortality estimates or conduct comparative mortality analyses for various ethnic groups because researchers have used denominator information from a census or survey and numerator data from death registration, sources that ascertain ethnicity in different ways (see Hahn & Stroup 1994; Rosenwaike & Bradshaw 1988; Sullivan et al. 1984). Some individuals, especially those who have resided in the U.S. for long periods of time, who are in exogenous marriages, or who are of mixed ethnicities, may be reported of one ethnicity in a census or survey and of another on death records (see Barringer, Gardner & Levin 1993). When inconsistencies arise, it is usually because individuals who were coded as

Mexican, Native, or Asian American in a census or survey were coded Caucasian at death (Sorlie, Rogot & Johnson 1992). This is especially apparent for Native Americans in California, Oklahoma, Oregon, and Washington (HHS 1993). These inconsistent codes can artificially lower life expectancy estimates for Caucasians and artificially raise them for other ethnic subpopulations. It seems desirable, therefore, that researchers make greater use of linked files, like the matched NHIS-NDI data set, where ethnic information is consistent for survivors and decedents. NCHS provides ethnicity on the NHIS-NDI file from the responses to the NHIS.

Because of the unique structure of this data set, we present some of the first national mortality comparisons of Native Americans to other ethnic groups. Mortality figures for Native Americans are infrequent; those figures that are presented are often for the Indian Health Service (IHS) areas, which include only the thirty-three states for which IHS is responsible (HHS 1993). These numbers are subject to substantial biases because of the different data sources for numerator and denominator data and different ways to ascertain ethnic identification. Thus, the Native American comparisons presented in this article are some of the very few national comparisons extant based on consistent ethnic codes for the numerator and denominator.

The NHIS asks whether the person falls into any of the following groups: Aleut, Eskimo, or American Indian; Asian or Pacific Islander; black; white; or other. A follow-up question asks whether the person or the person's ancestors is (are) Puerto Rican, Cuban, Mexican/Mexicano, Mexican American, Chicano, Other Latin American, or Other Spanish (Chyba & Washington 1993). In this study, we focus on the largest Hispanic subpopulation, Mexican Americans (Mexican/Mexicano, Mexican American, and Chicano). We exclude small groups of Cubans, Puerto Ricans, and other Hispanics because of insufficient numbers of deaths. To provide exclusive categories, we define our other categories as non-Hispanic Native Americans (Aleut, Eskimo, or American Indians), African Americans (blacks), Caucasian Americans (non-Hispanic whites), and Asian Americans (Asian or Pacific Islander). Although the Asian population is a diverse group in terms of country of origin, language, and customs, we examine the group in its entirety because NCHS does not differentiate Asian subpopulations and because there were few Asian deaths overall.[1] We follow convention by comparing the mortality of Mexican, Asian, Native, and African Americans to that of Caucasian Americans. The referent could, however, be any other ethnic group or, say, the average total U.S. population. The NHIS oversamples some ethnic groups, including Hispanics, to enable more accurate estimations.

Besides ethnicity, our independent variables are age, sex, education, income, marital status, and smoking status. Age is divided into five-year age categories except for the two extreme categories: under 20 (that is, ages 18 and 19), and ages 85 and above. We employed the linear relation of age because previous studies have demonstrated that age transformations of the five-year categories into age-logged or age-squared terms do not statistically improve the model fit (see, for examples, Elo & Preston 1994; Rogers 1992). We used four categories for education: eighth grade or less, ninth grade to eleventh grade, high school graduation (the referent), and some college (that is, thirteen years of education

or more). We include a broad measure of family income, measured as less than $20,000 or $20,000 or more for the following reasons. First, Williams and Collins (1995) report that "health gains due to income are small for households above $20,000 per year" (353). Second, the nonresponse rate is less than 2%, compared to a more detailed income measure, in which the nonresponse rate is 13%. Thus, we adopt the income measure with the highest response rate and that captures the major variations within health and mortality.[2] Marital status is coded as married (the referent), previously married, or never married. Smoking status is coded as never (the referent), former light (used to smoke less than a pack of cigarettes per day), former moderate-to-heavy (used to smoke a pack or more of cigarettes per day), current light (less than a pack of cigarettes per day), or current moderate-to-heavy (a pack or more of cigarettes per day). In those few instances that the marital status, income, education, or smoking values were missing, we stratified the sample by age, sex, and ethnicity, and imputed new values.

STATISTICAL ANALYSES

To examine ethnicity and mortality within a multivariate framework, we use discrete-time event history methods for categorical dependent variables (Allison 1982, 1984). The discrete-time hazard model, which is based on pooled intervals, deals with censored observations (deaths) and provides consistent estimators of the true standard errors (Teachman, Tedrow & Hill 1994). We recognize, of course, that the discrete-time model approximates hazard rates. The approximation is extremely good if the hazard of dying is relatively low (which it is until the highest ages). Moreover, even if the approximation is not extremely good, the values of the predictor variables will be similar. We report the coefficients in the form of odds ratios (see Hosmer & Lemeshow 1989). Because NHIS includes a complex clustering and stratification sampling design, we calculate hazard results for overall mortality with SUDAAN, a computer program that differentially weights sample strata to correctly estimate variances (Shah et al. 1990). For the cause-of-death analysis, we follow Allison's (1982) advice and use SUDAAN to calculate a series of equations, each of which compares individuals who die from a particular cause to those who survive the period. Because NHIS oversamples African Americans and Mexican Americans (see Massey et al. 1989), correctly incorporating the sample design generally produces larger standard errors and therefore more conservative statistical tests between ethnic groups.

We use five causes of death: four categories based on the three-digit codes presented in the *International Classification of Diseases* (HHS 1990) and a residual category. The four specific groupings are circulatory disease (390-459), cancer (ICD 140-239), respiratory disease (ICD 466-496), and social pathologies, which include suicide, accidents, homicide, and cirrhosis of the liver (ICD 571 and ICD E800-E999). This scheme is similar to those used in previous studies (i.e., Nam, Hummer & Rogers 1994; Potter 1991; Rogers 1992), although the categories are condensed because of cell sizes.

Results

Table 1 describes, by ethnicity, the demographic, familial, socioeconomic, and smoking status of those who were alive and those who died during the period from 1987 through 1991. For each ethnic group, individuals are generally more likely to die if they are older, male, previously married, less educated, or poorer, or if they smoked. Among survivors, there are variations in the distribution of these social and health characteristics by ethnic group. African Americans are the least likely to be married; a little over one-third of all African Americans are married. Asian Americans are by far the most educated ethnic group; more than 60% of Asians are college graduates. Asian and Caucasian Americans realize the greatest incomes; about 60% of each group earn over $20,000. Mexican Americans and Asian Americans tend not to smoke cigarettes; about 60% of Mexican Americans and more than 70% of Asians have never smoked cigarettes. Caucasians, on the other hand, have high rates of smoking, and especially high rates of quitting at high consumption levels. In fact, more than 15% of Caucasians are former heavy smokers, a status shown to be a high risk for several causes of mortality (Nam, Hummer & Rogers 1994).

Table 1 also reveals ethnic differences in causes of death. A high proportion of Asian Americans die from chronic degenerative diseases; circulatory diseases and cancer account for more than 80% of all Asian deaths. A substantial percentage of Caucasians die from these same causes, but many also die from respiratory diseases. Mexican, Native, and African Americans display high percentages of mortality from social pathologies. Indeed, more than one-quarter of all Native American deaths are attributed to cirrhosis of the liver and to accidents and other violent causes of death.

Table 1 includes sample sizes and shows some heterogeneity between ethnic groups and socioeconomic and behavioral factors. Of 398,204 total person-years, Caucasian Americans are the largest group, followed in turn by African, Mexican, Asian, and Native Americans. These distributions are affected, of course, by the focus on ages 18 and above and by the exclusion of other, smaller ethnic groups. Even so, Native Americans, while contributing almost 2,500 person-years, still represent a small subpopulation. Thus, because of relatively small sample sizes, estimates of Native and Asian American mortality should be interpreted cautiously. Some groups illustrate more homogeneity than others. For example, 85% of all Asian Americans have attained at least a high school diploma. Mexican Americans display much more educational variation: 50% have attained 12 years of schooling, but 30% have fewer than 9 years of education.

Table 1, while informative, does not simultaneously control for multiple factors. For such analyses, we turn to Table 2, which shows ethnic differences in mortality, controlling for various factors. By sequentially adding controls, we can determine the influence of each variable on the relationship between ethnicity and mortality. For example, a comparison of models 1 and 2 shows the effect of marital status on that relation. Comparing models 1 and 3 or models 1 and 4 shows the effect of socioeconomic status. Comparing models 1 and 6 shows how much smoking affects ethnic differences in mortality. Model 7 illustrates how all variables together affect ethnic mortality.

TABLE 1: Descriptive Statistics by Ethnicity, for Individuals 18 and Older Who Survived and Who Died: U.S., 1987-91[a]

	Mexican American		Native American		Asian American		African American		Caucasian American	
	Alive	Dead	Alive	Dead	Alive	Dead	Alive	Dead	Alive	Dead
Age										
Mean age	38.8	65.8	44.3	64.9	38.9	71.3	44.7	67.4	47.5	72.9
Sex										
Males	44.3	50.9	42.4	55.0	45.4	70.0	35.3	51.2	42.7	48.3
Females	55.7	49.1	57.6	45.0	54.6	30.0	64.7	48.8	57.3	51.7
Marital status										
Currently married	61.3	47.5	53.1	45.0	59.3	60.0	35.5	29.0	57.5	43.1
Previously married	18.3	49.1	27.9	45.0	11.1	30.0	35.0	56.8	24.8	49.6
Never married	20.4	3.4	19.1	10.0	29.6	10.0	29.7	14.2	17.7	7.3
Education										
8 years or less	30.5	67.8	16.2	50.0	8.5	45.0	15.8	46.8	8.6	30.3
9-11 years	17.3	11.9	18.0	20.0	6.6	15.0	20.2	19.2	11.0	17.0
12 years	31.9	11.9	38.9	15.0	22.5	20.0	37.2	22.3	38.5	30.9
13 or more years	20.3	8.4	26.9	15.0	62.4	20.0	26.7	12.7	41.8	21.8
Family income										
Less than $20,000	56.9	88.1	59.2	70.0	40.4	60.0	66.7	86.3	39.8	69.8
$20,000 or more	43.1	11.9	40.8	30.0	59.6	40.0	33.3	14.7	60.2	30.2
Smoking status										
Never	60.6	42.4	41.4	25.0	72.5	35.0	54.1	42.8	48.3	34.0
Former light	9.7	13.6	9.3	15.0	7.0	10.0	8.1	11.7	8.6	9.3
Former heavy	6.3	13.6	12.0	25.0	4.3	35.0	6.4	10.3	15.7	24.0
Current light	16.4	18.6	16.8	15.0	10.3	10.0	21.0	22.7	9.7	8.4
Current heavy	6.7	11.9	20.5	20.0	6.0	10.0	10.4	13.2	17.8	17.3
Cause of death										
Circulatory diseases		29.3		26.2		47.0		38.9		44.5
Cancer		28.0		27.3		34.6		23.6		30.1
Social pathologies		16.0		26.0		—‡		10.7		5.6
Respiratory diseases		4.4		—‡		—‡		5.4		6.8
Residual		22.3		18.3		12.6		21.4		13.0
Person-years		12,701		2,435		7,016		56,181		319,871

[a] Sources: Derived from NCHS 1992, 1993a, 1993b, 1994c. Age is given as the mean age; other figures are percentages. Percentages may not equal 100 due to rounding.

‡ Indicates fewer than 3 deaths

TABLE 2: Odds Ratios of Ethnic Differences in Overall Mortality, Controlling for Other Covariates: U.S., 1987-91[a]

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Ethnicity							
Mexican American	1.01	.99	.85	.93	.83	1.07	.90
Native American	1.02	1.01	.92	.98	.91	.97	.87
Asian American	.53†	.53†	.52*	.54†	.54†	.62	.62
African American	1.48***	1.42***	1.32***	1.35***	1.23**	1.51***	1.28***
Caucasian American	ref	ref	ref	ref	ref	ref	ref
Demographic status							
Age	1.53***	1.52***	1.49***	1.49***	1.47***	1.57***	1.51***
Sex	1.78***	1.95***	1.79***	1.85***	1.96***	1.57***	1.68***
Marital status							
Currently married		ref			ref		ref
Previously married		1.32***			1.24***		1.21***
Never married		1.40***			1.32***		1.42***
Socioeconomic status							
Education							
8 years or less			1.34***		1.24***		1.24***
9-11 years			1.22***		1.16*		1.12*
12 years			ref		ref		ref
13 or more years			.75***		.78***		.81***
Income				1.51***	1.30***		1.28***
Smoking status							
Never smoked						ref	ref
Former light						1.16*	1.20**
Former heavy						1.41***	1.46***
Current light						1.88***	1.83***
Current heavy						2.33***	2.24***
Log likelihood	19,573.02	19,538.98	19,492.81	19,503.83	19,439.11	19,408.02	19,291.13

[a] Sources: Derived from NCHS 1992, 1993a, 1993b, 1994c.

† p ≤ .10 * p ≤ .05 ** p ≤ .01 *** p ≤ .001

Model 1 incorporates a conventional demographic approach to ethnic differences in mortality by controlling for age and sex. It shows that, compared to Caucasian Americans, African, Mexican, and Native Americans have higher mortality and Asian Americans exhibit lower mortality. Compared to Caucasian Americans, African Americans are about 50% more likely to die, whereas Asians are almost half as likely to die. The mortality of Mexican and Native Americans is similar to the mortality of Caucasian Americans. The subsequent models show whether these general patterns change with the introduction of other social, economic, or behavioral health factors.

Model 2, which controls for marital status, does not substantially alter ethnic mortality patterns. But models 3, 4, and 5, which control for socioeconomic differences, show that socioeconomic status substantially affects the mortality gap between Caucasian and Mexican, Native, and African Americans. For instance, controlling for age and sex reveals a 48% gap between African and Caucasian Americans, but simultaneously controlling for education reduces the gap to a 32% difference. Compared to Caucasian Americans, Mexican Americans exhibit the same odds of death when age and sex are controlled, but lower odds when education is also controlled. Thus, a major contributor to ethnic differences in mortality is socioeconomic status. The fact that Mexican, Native, and African Americans acquire less education and income than Caucasian Americans contributes to their mortality.

Controlling for smoking status, in model 6, substantially affects ethnic differences in overall mortality for Mexican, Asian, and Caucasian Americans. Mexican and Caucasian Americans exhibit the same risk of dying in model 1, but once smoking is controlled, Mexican Americans exhibit a higher risk of death. Thus, once the risky smoking behavior of Caucasians is taken into account, the mortality gap between them and Mexican Americans widens. Through similar logic, the statistically significant differences between Asian and Caucasian mortality in model 1 disappears with controls for smoking. Thus, the high rate of smoking among Caucasian Americans hampers their ability to mirror the low Asian mortality and the low smoking prevalence rates among Mexican Americans helps compensate for their lower socioeconomic status and provides them with mortality rates comparable to their Caucasian counterparts.

Model 7 simultaneously controls for demographic, social, and behavioral health characteristics. With such controls, there is no significant mortality difference between Caucasian and Mexican Americans or between Caucasian and Native Americans, the gap narrows substantially between Caucasian and African Americans and also narrows between Caucasian and Asian Americans. Other referents than Caucasians would of course yield different odds ratios. For instance, model 7 (including all variables) shows that African Americans are 28% more likely to die than Caucasian Americans; but over twice as likely to die as Asians $\left(e^{(\ln(1.28)-\ln(.62))}\right)$. This table underscores the impact of familial, socioeconomic, and behavioral health variables on ethnic differences in overall mortality. To understand which causes of death contribute to these mortality differences, we turn to Table 3.

Table 3 shows cause-specific ethnic differences in mortality, controlling for demographic, social, and economic factors and for smoking. Compared to Caucasian Americans, African Americans demonstrate slightly higher mortality from cancer and circulatory diseases and significantly higher mortality from social pathologies and residual causes; Mexican and Native Americans also have higher mortality rates from social pathologies. In fact, Native Americans have over 2.7 times the Caucasian American mortality rate in this category, which is dominated by automobile accidents. Asian Americans have low mortality from all causes when compared to Caucasian Americans. To determine which factors are most responsible for these ethnic patterns in mortality, we turn to Table 4.

Table 4 presents four separate models for cause-specific mortality: one controlling for demographic variables only, one controlling for demographic and

TABLE 3: Odds Ratios of Ethnic Differences in Cause-Specific Mortality, Controlling for Other Covariates: U.S., 1987-91[a]

	Cancer	Circulatory Diseases	Respiratory Diseases	Social Pathologies	Residual
Ethnicity					
Mexican American	.91	.66	.60	1.23	.76
Native American	.76	.57	‡	2.74†	.78
Asian American	.77	.73	‡	‡	.41†
African American	1.13	1.16	.88	1.65**	2.08***
Caucasian American	ref	ref	ref	ref	ref
Demographic status					
Age	1.53***	1.66**	1.60***	1.12***	1.52***
Sex	1.35***	1.58**	2.09***	4.04***	1.77***
Marital status					
Currently married	ref	ref	ref	ref	ref
Previously married	.92	1.09	1.99***	1.95***	1.34*
Never married	.93	1.18	1.63	1.76**	1.94**
Socioeconomic status					
Education					
8 years or less	1.15	1.20*	1.40†	1.60*	1.01
9-11 years	1.05	1.09	1.31	1.46†	.91
12 years	ref	ref	ref	ref	ref
13 or more years	.73**	.79**	.82	.89	1.23
Income	1.10	1.35***	1.79**	1.43†	1.05
Smoking status					
Never smoked	ref	ref	ref	ref	ref
Former light	1.58**	1.06	2.38**	.40**	1.39
Former heavy	2.07***	1.33***	4.18***	.69	2.15***
Current light	2.39***	1.70***	6.57***	1.33	1.00
Current heavy	3.78***	1.92***	4.57***	1.27	1.01†

[a] Sources: Derived from NCHS 1992, 1993a, 1993b, 1994c.
‡ Indicates fewer than 3 deaths.

† $p \leq .10$ * $p \leq .05$ ** $p \leq .01$ *** $p \leq .001$

TABLE 4: Odds Ratios of Ethnic Differences in Cause-Specific Mortality with and without Controls for Demographic, Social, and Health Covariates: U.S., 1987-91

	Circulatory Diseases				Cancer			
Ethnicity	Demo	Social	Health	Full	Demo	Social	Health	Full
Mexican American	.75	.63	.78	.66	.89	.78	1.03	.91
Native American	.64	.58	.63	.57	.88	.81	.82	.76
Asian American	.65	.66	.72	.73	.58	.59	.77	.77
African American	1.35***	1.14	1.36***	1.16	1.15	1.03	1.23†	1.13
Caucasian American	ref	ref	ref	ref	ref	ref	ref	ref

	Respiratory Diseases			
Ethnicity	Demo	Social	Health	Full
Mexican American	.70	.53	.78	.60
Native American	—‡	—‡	—‡	—‡
Asian American	—‡	—‡	—‡	—‡
African American	1.24	.86	1.23	.88
Caucasian American	ref	ref	ref	ref

	Social Pathologies				Residual			
Ethnicity	Demo	Social	Health	Full	Demo	Social	Health	Full
Mexican American	1.62	1.21	1.60	1.23	.77	.71	.81	.76
Native American	3.58*	2.88*	3.38*	2.74†	.95	.86	.86	.78
Asian American	—‡	—‡	—‡	—‡	.37*	.37*	.41†	.41†
African American	2.39***	1.69**	2.23***	1.65**	2.22***	2.00***	2.25***	2.08***
Caucasian American	ref	ref	ref	ref	ref	ref	ref	ref

[a] *Demo* indicates mortality differences in the basic demographic models, with controls only for age and sex. *Social* indicates mortality differences after controls for age, sex, marital status, education, and income. *Health* indicates mortality differences after controls for age, sex, and smoking status. *Full* indicates mortality differences after controls for age, sex, marital status, education, income, and smoking status.

Sources: Full model derived from Table 3. Other models derived from NCHS 1992, 1993a, 1993b, 1994c.

—‡ Indicates fewer than 3 deaths.

† $p \leq .10$ * $p \leq .05$ ** $p \leq .01$ *** $p \leq .001$

social factors, another controlling for demographic factors and smoking, and last, a replication of the full model used in Table 3. Compared to Caucasian American mortality, Mexican, Native, and African American mortality is reduced with controls for socioeconomic status. Such reductions are small for cancer and circulatory diseases, where the mortality gap between Caucasians and other groups is already small, but are larger for social pathologies. For example, compared to Caucasian Americans of the same age and sex, Native Americans are 3.6 times as likely to die from social pathologies but only 2.9 times as likely to die if we control for social factors. Similar results are obtained for Mexican and African Americans. Thus, the high mortality rate from social pathologies among Mexican, Native, and African Americans is due in large part to social characteristics — differences in marital status, education, and income.

Compared to Caucasian Americans of the same age and sex, African Americans are 35% more likely to die from circulatory diseases. But when socioeconomic status and marital status are controlled, the gap closes significantly, to a 14% difference. The gap between Caucasian and African Americans witnesses similar narrowing once socioeconomic status is controlled for all other causes of death.

Compared to Caucasian Americans of the same age and sex, Mexican Americans experience the same or lower mortality for circulatory diseases, cancer, respiratory diseases, and residual causes of death. Once socioeconomic status and marital status are controlled, Mexican American mortality drops further for all these causes of death. But when cigarette smoking is controlled, Mexican American mortality increases — for cancer, by 14%. Similar relations exist for African Americans: whereas African Americans are just 15% more likely than Caucasian Americans to die from cancer in the demographic model, they are almost 25% more likely to die when cigarette smoking is controlled. Because Caucasian Americans exhibit heavy rates of cigarette consumption, followed by African Americans and then by Mexican Americans, controlling for smoking widens the mortality gap for Caucasian and African Americans and for Caucasian and Mexican Americans for smoking-related causes of death.

On the other hand, controlling for cigarette smoking generally narrows the cause-specific mortality gaps between Asian and Caucasian Americans. Because Caucasians smoke more than Asian Americans, and because Asians enjoy lower mortality in general, controlling for smoking reduces the gap between these two groups for residual causes of death and for the two causes of death directly affected by smoking: circulatory diseases and cancer. Asians are 42% less likely than Caucasians to die from cancer when age and sex are controlled, but only 23% less likely than Caucasians to die when cigarette smoking is controlled.

Conclusions

This study has provided new, detailed, and comparative mortality results by ethnicity. Past studies have often compared only one or two ethnic groups, have used disparate categorization between numerator and denominator data, and have not controlled for other covariates. Therefore, our results offer more

detailed, definitive results by ethnic groups on a national level and provide controls for covariates.

This study underscores the importance of examining ethnic differences from social and behavioral health as well as demographic perspectives. For example, from a social characteristics perspective, we find that the high mortality from accidents and violence for Mexican, Native, and African Americans is due in large part to higher rates of poverty, lower levels of education, and marital disruption — that is, to their social disadvantages. The behavioral health characteristics perspective helps to explain the epidemiologic paradox, or why Mexican and Caucasian Americans experience similar mortality rates, even though a greater proportion of Mexican Americans live in poverty and lack health insurance (Markides & Coreil 1986; Sorlie et al. 1993). Our findings suggest that the healthy behavior of Mexican Americans (their tendency not to smoke) compensates in part for their social disadvantages. In contrast, Caucasians as a group compromise their low mortality through heavy, prolonged smoking.

We followed convention by using Caucasians as the referent group. An alternate approach would be to use Asians, the group with the lowest mortality. The tables provide the information needed for such comparisons, which would increase the odds ratio for many ethnic groups; as we have already demonstrated from Table 2, African Americans would move from an overall mortality excess of 28% (compared to Caucasians) to an excess greater than twofold (compared to Asians).

The demographic, social characteristic, and behavioral perspectives each aid in the explanation of ethnic differences in mortality. For instance, the sources of mortality differences are more clearly understood when demographic characteristics are taken into account. However, it is evident from this and other research that ethnic mortality differences exist beyond the demographic composition of the respective populations.

This article also demonstrates that the social characteristic hypothesis is important in explaining ethnic mortality differences. Asian Americans clearly have the lowest mortality and most advantaged socioeconomic composition of the groups studied; when socioeconomic characteristics are controlled, their mortality advantage relative to the other groups decreases. Similarly, the least socioeconomically advantaged groups — Mexican, Native, and African Americans — are able to reduce their mortality gaps with Caucasian and Asian Americans when social characteristics are controlled. Thus, improvements in further reducing socioeconomic gaps between U.S. ethnic groups may also have an impact in reducing mortality differences between these groups. Furthermore, the impact of the social characteristic explanation may even be stronger if additional measures of social inequality — such as wealth, occupational prestige, and discrimination — were measured and taken into account (Hummer 1996).

In addition to the demographic and social characteristic perspectives, this article demonstrates that the behavioral health perspective helps in understanding ethnic mortality differences. However, controlling for such factors — in this case smoking — may also work to statistically widen mortality differences. Here, it is clear that the mortality advantage of Caucasian Americans relative to

African, Mexican, and Native Americans would be even greater if cigarette smoking were eliminated. It would be interesting to speculate whether similar effects would be observed if other health behaviors, such as diet or the use of alcohol or drugs, were examined.

The NHIS-NDI linked data set enables researchers to better examine ethnic mortality patterns. Future research could explore multiple-cause mortality, which considers all the identifiable medical causes entered on the death certificate (see Nam 1990). Combinations of causes can be more lethal than single causes, can affect the timing of death, and can influence the mortality differences between subpopulations.

It is also important to examine mortality differences within ethnic subpopulations. Some groups used here have subgroups that vary substantially by history, geography, language, and custom; consider the differences among Apaches, Navahos, and Sioux or among Chinese, Japanese, and Laotians. Researchers could begin examining different African American subpopulations, especially as larger numbers of individuals immigrate to the U.S. from Africa, the West Indies, South and Central America, and Europe (Bryce-Laporte 1993). And although Caucasian Americans have conventionally been treated as a homogeneous group, there are differences by country of origin, language spoken, religion, and time since migration that may influence mortality (Rosenwaike & Hempstead 1990).

Mortality is a driving force behind the size and composition of ethnic populations, which in turn affect natural increase, age structure, cohort effects, and family formation and composition. Therefore, information on the current and future mortality of ethnic groups is crucial to understanding the dynamics of our society. And these results give us added insight into the demographic, social, and health mechanisms that allow us to persevere or to perish.

Notes

1. The U.S. Bureau of the Census is primarily responsible for the design and collection of NHIS data. The NHIS sample includes about 10% of the approximately 1,900 primary sampling units (PSUs) that cover the U.S. "A PSU consists of a county, a small group of contiguous counties, or a metropolitan statistical area" (Massey et al. 1989:2). Although NHIS data represent the entire U.S. population, sampling by geographic areas means that larger subgroups within the Asian and Native American populations are more likely to represent the entire ethnic population. For example, mortality rates for Native Americans may reflect more the Cherokee and Navaho experiences, the two largest Indian subpopulations, than, say, those of the Acoma or Taos Indians. Similarly, mortality rates for Asians may reflect more the Chinese experience, the largest U.S. Asian subpopulation, than, say, that of the Hmong (U.S. Bureau of the Census 1994). If the mortality experiences of the Hmong and Chinese are similar, then no biases are introduced. If the Hmong have much higher mortality than the Chinese, then our Asian estimates will be artificially low. But such bias will be slight because even high mortality contributed by a proportionately small group will negligibly affect the overall ethnic group's mortality.

2. Different approaches toward capturing income differences may yield slightly different results and interpretations. For other income adjustments, refer to Buhmann et al. (1988) for income equivalence scales, or Ruggles (1990) for poverty indexes. Although we assert that mortality is reflected more directly and accurately by an income measure that is answered by the greatest proportion of respondents, that is clearly constructed, and that is not contaminated by other variables, we leave this issue to the reader to weigh and for future research to resolve.

References

Allison, Paul D. 1982. "Discrete-Time Methods for the Analysis of Event Histories." *Sociological Methodology* 12:61-98.

____. 1984. *Event History Analysis: Regression for Longitudinal Event Data.* Sage University Press.

Barringer, Herbert R., Robert W. Gardner, and Michael J. Levin. 1993. *Asians and Pacific Islanders in the United States.* Russell Sage Foundation.

Bryce-Laporte, Roy Simon. 1993. "Voluntary Immigration and Continuing Encounters between Blacks: The Post-Quincentenary Challenge." *Annals of the American Academy of Political and Social Science* 530:28-41.

Buhmann, Brigitte, Lee Rainwater, Guenther Schmaus, and Timothy M. Smeeding. 1988. "Equivalence Scales, Well-Being, Inequality, and Poverty: Sensitivity Estimates across Ten Countries Using the Luxembourg Income (LIS) Database." *Review of Income and Wealth* 34:115-20.

Castro, Felipe G., Lourdes Baezconde-Garbanati, and Hector Beltran. 1985. "Risk Factors for Coronary Heart Disease in Hispanic Populations: A Review." *Hispanic Journal of Behavioral Sciences* 7:153-75.

Centers for Disease Control. 1989. *Smoking Tobacco and Health: A Fact Book.* U.S. Department of Health and Human Services. DHHS Publication no. (CDC) 89-8397.

Centerwall, Brandon S. 1984. "Race, Socioeconomic Status, and Domestic Homicide, Atlanta, 1971-72." *American Journal of Public Health* 74:813-15.

Chyba, Michele M., and Linda R. Washington. 1993. "Questionnaires from the National Health Interview Survey, 1985-89." *Vital and Health Statistics* 1:1-412.

Cooper, Richard. 1991. "Celebrate Diversity — Or Should We?" *Ethnicity and Disease* 1:3-8.

____. 1994. "A Case Study in the Use of Race and Ethnicity in Public Health Surveillance." *Public Health Reports* 109:46-52.

Crews, Douglas, and James Bindon. 1991. "Ethnicity as a Taxonomic Tool in Biomedical and Biosocial Research." *Ethnicity and Disease* 1:42-49.

Dressler, William. 1993. "Health in the African American Community: Accounting for Health Differentials." *Medical Anthropology Quarterly* 7:325-45.

Elo, Irma T., and Samuel H. Preston. 1994. "Educational Differentials in Mortality: United States, 1979-85." Paper presented to the annual meeting of the Population Association of America, Miami, Florida.

Fingerhut, Lois A., Cheryl Jones, and Diane M. Makuc. 1994. "Firearm and Motor Vehicle Injury Mortality — Variations by State, Race, and Ethnicity: United States, 1990-91. *Advance Data from Vital and Health Statistics* 242:1-12.

Frisbie, W. Parker, and Frank D. Bean. 1978. "Some Issues in the Demographic Study of Racial and Ethnic Populations." Pp. 1-11 in *The Demography of Racial and Ethnic Groups,* edited by Frank D. Bean and W. Parker Frisbie. Academic Press.

Hahn, Robert A., and Donna F. Stroup. 1994. "Race and Ethnicity in Public Health Surveillance: Criteria for the Scientific Use of Social Categories." *Public Health Reports* 109:7-15.

Horm, John. 1993. "The National Health Interview Survey and the National Death Index." Presentation to the National Center for Health Statistics on 19 May, Hyattsville, Maryland.

Hosmer, David W., and Stanley Lemeshow. 1989. *Applied Logistic Regression.* John Wiley & Sons.

Hummer, Robert A. 1996. "Black-White Differences in Health and Mortality: A Review and Conceptual Model." *Sociological Quarterly* 37:105-25.

Jaynes, Gerald D, and Robin M. Williams Jr. (eds.). 1989. *A Common Destiny: Blacks and American Society.* National Academy Press.

Kabat, Geoffrey C., Alfredo Morabia, and Ernst L. Wynder. 1991. "Comparison of Smoking Habits of Blacks and Whites in a Case-Control Study." *American Journal of Public Health* 81:1483-86.

Kochanek, Kenneth D., Jeffery D. Maurer, and Harry M. Rosenberg. 1994. "Why Did Black Life Expectancy Decline from 1984 through 1989 in the United States?" *American Journal of Public Health* 84:938-44.

Loya, Fred, and James A. Mercy. 1985. *The Epidemiology of Homicide in the City of Los Angeles 1970-1979*. Department of Health and Human Services, Public Health Service, Centers for Disease Control.

Manton, Kenneth G., Clifford H. Patrick, and Katrina W. Johnson. 1987. "Health Differentials between Blacks and Whites: Recent Trends in Mortality and Morbidity." *Milbank Quarterly* 65(suppl. 1):129-99.

Markides, Kyriakos S., and Jeannie Coreil. 1986. "The Health of Hispanics in the Southwestern United States: An Epidemiological Paradox." *Public Health Reports* 101:253-65.

Massey, James T., Thomas F. Moore, Van L. Parsons, William Tadros. 1989. "Design and Estimation for the National Health Interview Survey, 1985-94." *Vital and Health Statistics* 2:1-33.

Mosley, W. Henry, and Lincoln C. Chen. 1984. "An Analytic Framework for the Study of Child Survival in Developing Countries." *Population and Development Review* 10(suppl.):25-45.

Nam, Charles B. 1990. "Mortality Differentials from a Multiple Cause-of-Death Perspective." Pp. 328-42 in *Measurement and Analysis of Mortality*, edited by Jacques Vallin, Stan D'Souza, and Alberto Palloni. Oxford University Press.

____. 1995. "Another Look at Mortality Crossovers." *Social Biology* 41:133-42.

Nam, Charles B., Robert A. Hummer, and Richard G. Rogers. 1994. "Underlying and Multiple Causes of Death Related to Smoking." *Population Research and Policy Review* 13:305-25.

National Center for Health Statistics. 1980. DHEW Publication no. (PHS) 80-1247. NCHS.

____. 1992. *1988 National Health Interview Survey*. CD-ROM Series 10, no. 2. NCHS.

____. 1993a. *1987 National Health Interview Survey*. CD-ROM Series 10, no. 1. NCHS.

____. 1993b. *1990 National Health Interview Survey*. CD-ROM Series 10, no. 4. NCHS.

____. 1994a. "Advance Report of Final Mortality Statistics, 1992." *Monthly Vital Statistics Report* 43:1-76.

____. 1994b. *Health, United States, 1993*. Public Health Service.

____. 1994c. *1986 National Health Interview Survey: Multiple Cause of Death Public Use Data.* Diskette and documentation. NCHS.

Patterson, Blossom H., and Robert Bilgrad. 1986. "Use of the National Death Index in Cancer Studies." *Journal of the National Cancer Institute* 77:877-81.

Potter, Lloyd B. 1991. "Socioeconomic Determinants of White-Black Male Life Expectancy." *Demography* 28:303-21.

Rogers, Richard G. 1992. "Living and Dying in the USA: Sociodemographic Determinants of Death among Blacks and Whites." *Demography* 29:287-303.

____. 1996. "Mortality Differentials in a Diverse Society." Forthcoming in *American Diversity: A Demographic Challenge for the Twenty-First Century*, edited by Nancy A. Denton and Stewart E. Tolnay. SUNY Press.

Rosenwaike, Ira. 1987. "Mortality Differentials among Persons Born in Cuba, Mexico, and Puerto Rico Residing in the United States, 1979-81." *American Journal of Public Health* 77:603-6.

____. 1988. "Cancer Mortality among Mexican Immigrants in the United States." *Public Health Reports* 103:195-201.

Rosenwaike, Ira, and Benjamin S. Bradshaw. 1988. "The Status of Death Statistics for the Hispanic Population of the Southwest." *Social Science Quarterly* 69:722-36.

Rosenwaike, Ira, and Katherine Hempstead. 1990. "Differential Mortality by Ethnicity and Nativity: Foreign- and Native-Born Irish, Italians, and Jews in New York City, 1979-1981." *Social Biology* 37:11-25.

Ruggles, Patricia. 1990. *Drawing the Line: Alternative Poverty Measures and Their Implication for Public Policy*. Urban Institute Press.

Schoenborn, Charlotte A. 1986. "Health Habits of U.S. Adults, 1985: The 'Alameda 7' Revisited." *Public Health Reports* 101:571-80.

Shah, Babubhai V., Beth G. Barnwell, P. Nileen Hunt, and Lisa M. LaVange. 1990. *SUDAAN: Software for SUrvey DAta ANalysis*. Research Triangle Institute.

Shai, Donna, Ira Rosenwaike, and Richard G. Rogers. 1991. "Mortality by Violence among Mexican Immigrants and Mexican-Americans in California and Texas." Pp. 161-67 in *Mortality of Hispanic Populations*, edited by Ira Rosenwaike. Greenwood Press.

Simpson, Sylvia G., Raymond Reid, Susan P. Baker, and Stephen Teret. 1983. "Injuries among the Hopi Indians: A Population-Based Survey." *Journal of the American Medical Association* 249:1873-76.

Snipp, C. Matthew. 1989. *American Indians: The First of This Land*. Russell Sage Foundation.

Sorlie, Paul D., Eric Backlund, Norman J. Johnson, and Eugene Rogot. 1993. "Mortality by Hispanic Status in the United States." *Journal of the American Medical Association* 270:2464-68.

Sorlie, Paul, Eugene Rogot, Roger Anderson, Norman J. Johnson, and Eric Backlund. 1992. "Black-White Mortality Differences by Family Income." *Lancet* 340:346-50.

Sorlie, Paul D., Eugene Rogot, and Norman J. Johnson. 1992. "Validity of Demographic Characteristics on the Death Certificate." *Epidemiology* 3:181-84.

Sterling, Theodor, Wilf Rosenbaum, and James Weinkam. 1993. "Income, Race, and Mortality." *Journal of the National Medical Association* 85:906-11.

Sullivan, Teresa A., Francis P. Gillespie, Michael Hout, and Richard G. Rogers. 1984. "Alternative Estimates of Mexican American Mortality in Texas, 1980." *Social Science Quarterly* 65:609-17.

Teachman, Jay D., Lucky Tedrow, and Daniel Hill. 1994. "A Note on Discrete Time Hazard Rate Models: Estimated Parameters and Standard Errors." Mimeo.

U.S. Bureau of the Census. 1994. *Statistical Abstract of the United States*. Government Printing Office.

U.S. Department of Health and Human Services (HHS). 1990. *International Classification of Diseases*. Vol. 1, 9th Rev., 3d ed. DHHS Publication (PHS) 89-1260. Government Printing Office.

____. *Trends in Indian Health*. 1993. Government Printing Office.

Warshauer, M. Ellen, and Mary Monk. 1978. "Problems in Suicide Statistics for Whites and Blacks." *American Journal of Public Health* 68:383-88.

Williams, David. 1990. "Socioeconomic Differences in Health: A Review and Redirection." *Social Psychology Quarterly* 53:81-99.

Williams, David, and Chiquita Collins. 1995. "U.S. Socioeconomic and Racial Differences in Health: Patterns and Explanations." Pp. 349-86 in *Annual Review of Sociology*. Vol. 21. Annual Reviews.

Williams, David, Risa Lavizzo-Mourey, and Rueben C. Warren. 1994. "The Concept of Race and Health Status in America." *Public Health Reports* 109:26-41.

Zane, Nolan, and Jeannie Huh Kim. 1994. "Substance Use and Abuse." Pp. 316-43 in *Confronting Critical Health Issues of Asian and Pacific Islander Americans*, edited by Nolan Zane, David Takeuchi, and Kathleen Young. Sage.

Book Reviews

Mental Disorder and Crime.
Edited by Sheilagh Hodgins. Sage Publications. 1993, 339 pp. Cloth, $55.00; paper, $26.00.

Reviewer: WALTER R. GOVE, *Vanderbilt University*

According to Sheilagh Hodgins people frequently fear that the mentally ill are likely to be violent, but members of the scientific community have not been convinced this is the case. However, she holds that "in recent years much evidence has suggested that major mental disorders (schizophrenic, major affective disorders, and other psychiatric disorders) and violence are associated," and this provides the premise around which this book is framed. The book is divided into three parts. Part 1, "The Criminology of Mentally Disordered Persons," is comprised of six papers that are largely concerned with developing a descriptive understanding of the association between mental disorders, crime, and violent behavior. Part 2, "Risk Factors for Violent Behavior," contains eight papers, most of which are concerned with developing an understanding of the causes of violence. The concluding part of the book has three papers that provide a rough assessment of the evidence linking mental disorders and crime and discusses implications for intervention.

This book is an attempt to present an overview of recent evidence on the relationship between mental disorders and crime and to indicate what is unknown but needs to be known. As a consequence of this latter concern, a few of the papers are primarily a description of, and justification for, major research programs presently in process. To fully understand and evaluate the papers dealing with the relationship between mental disorders and crime the reader needs a working knowledge of the Diagnostic and Statistical Manual (DSM III, DSMIIIR, or DSMIV) and how the mental disorders described in this manual are measured in community studies (e.g., the Diagnostic Interview schedule [DIS] used in the Epidemiologic Attachment Area [ECA] study). This is particularly important as the term *mental disorder* is used differently by the various authors; fortunately, how it is being used is clearly stated at the beginning of each chapter. To fully understand and evaluate many of the papers in part 2 the reader will need an understanding of basic knowledge in neuroscience and behavioral endocrinology.

This book is aimed at a professional audience and all the chapters have something to recommend them. I personally learned the most from the following chapters: (1) Sheilagh Hodgins on "The Criminality of Mentally Disordered Persons," which includes detailed data on how the comorbidity of mental disorders is related to crime in the Swedish Metropolitan Project; (2) Lee Robins, who makes a similar, but much more elaborate presentation using the ECA data;

(3) Pamela Taylor who shows how the violent criminal behavior of persons who have an acute psychotic disorder differs from most violent criminal behavior; (4) Matti Virkkunen and Merku Linnoila, who indicate the ways in which the neurophysiology of the brain, particularly low levels of serotonin, are related to violent behavior; (5) Robert 0. Pihl and Jordan Peterson, who describe the relationship between alcohol and aggressive behavior and focus on how alcohol may *increase* pain sensitivity while simultaneously reducing both the fear associated with potential threats and a person's response flexibility; and (6) John Monchan, who gives a good overview of the evidence between mental disorders and violence.

Taken as a whole I think the material in this book lends itself to a number of conclusions that can be reached with a reasonably high level of certainty: (1) Persons who abuse alcohol and/or drugs have very high rates of criminal behavior (chapters 1, 9, and 15). (2) Persons with a psychopathological (or antisocial or sociopathic) personality have very high rates of criminal behavior (chapters 6, 7, 9, 11, 12, and 15). Furthermore, the relationship between psychopathic traits is not substantively affected by controls for serious mental disorders such as psychosis (chapter 6). (3) Persons with major disorders of thought or affect (e.g. psychoses, schizophrenia, affective psychoses, paranoid states) have relatively high levels of criminality (chapters 1, 9, and 15). However, when substance abuse (chapter 1) or substance abuse and an antisocial behavior (chapter 9) are controlled for, the relationship between the major mental disorders and the high rate of criminality becomes statistically nonsignificant. (4) Most violent psychotic offenders differ from other violent offenders in that their violent behavior appears to be a direct response to psychotic symptoms, particularly delusions and hallucinations (chapter 4, also see chapters 3 and 15). (5) Disruptive parenting practices result in antisocial behavior in children and are a precursor of aggressive offending in personality disordered adults (chapters 7 and 8). (6) Understanding why the variation among persons in impulsivity, hyperactivity, and sensation seeking are linked to habitual violent behavior requires an understanding of basic neuroanatomy, and the ways in which monoamine oxidase activity, serotonin, and other neurotransmitters affect the functioning of the brain (chapters 10, 11, and 12). As the work of sociologists increasingly indicates that impulsivity and sensation seeking is related to criminal activity, sociologists need to be taking this literature into account.

This is an important book. However, I am bothered by the fact that no one raises the possibility that the criminal behavior of persons with personality disorders and persons who abuse alcohol and/or drugs might be best viewed as volitional acts of deviant behavior. As this issue is not addressed, the book is indicative of a trend to medicalize deviant behavior without a consideration of all the issues involved.

Cultivating Differences: Symbolic Boundaries and the Making of Inequality. *Edited by Michèle Lamont and Marcel Fournier.* University of Chicago Press, 1992. 346 pp. Cloth, $49.95; paper, $20.00.

Reviewer: ANNE BOWLER, *University of Delaware*

This excellent volume is required reading for sociologists working in the area of culture and the arts. It also should be of interest to sociologists seeking a clearer understanding of the role of culture in "the making of inequality." *Cultivating Differences* is distinguished by the quality of its individual contributions, whose authors bring a sophisticated approach to the subject of culture, including Herbert Gans in his wonderful preface. This does not mean that readers should expect to find a single theoretical orientation or model. Rather, each contributor rejects the generally reductive conclusion that culture can be read as a simple reflection of social structure and instead recognizes and seeks to explicate the complexity of the culture-society relationship.

Part 1 explores the institutionalization of cultural categories in modern society. Paul DiMaggio maps the establishment of a "high cultural model" in American drama, opera, and dance in the first four decades of this century. Diana Crane challenges the conceptual adequacy of the categories of high and popular culture in contemporary society. Arguing that "it is more meaningful to categorize recorded cultures in terms of the environments in which they are created, produced, and disseminated," she offers insight into conditions under which institutional arrangements shape systems of classification. Joseph Gusfield analyzes the discourse of purity and the body in the history of the natural foods movement. Nicola Beisel examines the moral crusade led by American reform leader Anthony Comstock, who mobilized the social categories of youth, class, and ethnicity to redefine boundaries between obscenity and literature in the late nineteenth century.

Part 3 reexamines a theme of earlier work in the sociology of culture: the role of high culture as a tool for creating and reinforcing inequalities. David Halle challenges the validity of this function in his study of consumers of abstract art, reporting on buyers who eschew claims to the superior intellectual capacities required to admire abstract art in favor of the decorative qualities of works. In contrast, essays by Richard Peterson and Albert Simkus on musical taste and Vera Zolberg on the museum indicate that high culture continues to be an important signifier of status, albeit in increasingly more complex and variegated ways. Peterson and Simkus examine the relationship between musical taste and occupational status. While they find a hierarchy of cultural preferences by class, it is "pyramidal" in structure, with less consensus on ranking as one moves down the social scale. In a comparative analysis of French and American art museums, Zolberg explains why the democratic goals of museum administrators often remain unfulfilled.

Part 3 demonstrates the significance of gender and ethnicity as bases of distinction and stratification. Randall Collins maps the intersection of class and gender in the "status-producing" activities performed by women in the workplace and at home. Cynthia Fuchs Epstein elucidates the cultural bases and dynamics of gender boundaries in the workplace. In an essay directed in part toward a

retheorization of Bourdieu's model of cultural capital, John Hall argues that neither gender nor ethnicity can be viewed as "secondary" to class.

Finally, part 4 raises the question of exclusion and the polity. Jeffrey Alexander examines the symbolic dimension of civil society. He presents a critical exegesis of the binary character of "civil discourse" as it operates at three levels (motives, relations, institutions), dividing society into "those who deserve inclusion and those who do not." Alan Wolfe, in turn, locates a tension between the very terms *inclusion* and *exclusion* in a liberal democratic society. The history of modern democracy is one of inclusion and yet a totally inclusive society, Wolfe argues, is neither possible nor desirable. The key for Wolfe is that not all boundaries are hierarchical and oppressive nor do all distinctions involve a process of "othering." The task then becomes one that combines a democratic commitment to universal rights with respect for specific differences.

The Dream of a New Social Order: Popular Magazines in America, 1893-1914.
By Matthew Schneirov. Columbia University Press, 1994. 357 pp. $29.50.

Reviewer: KAREN A. CERULO, *Rutgers University*

Matthew Schneirov's *The Dream of a New Social Order* provides a fascinating window on one of the most pivotal periods in U.S. history, the decades that bridged the nineteenth and twentieth century. For many that window may seemingly emanate from an unexpected source, for it is through the pages of popular magazines that Schneirov provides this slice of social history.

Schneirov guides his readers through the transition from homogeneous to heterogeneous America. In 1893 the U.S. was still dominated by middle- and upper-class Protestant ideals, but by 1914 the nation was grappling with an extended set of beliefs and practices — values and norms generated by new social phenomena such as the growth of cities, diverse immigration, industrial expansion, and sheer social energy. Popular magazines both reflected and, according to Schneirov, provoked the changing tenor of the times. Indeed, Schneirov convincingly demonstrates the ways in which magazine editors of the period intentionally blurred the boundaries and categories of postbellum America in an effective effort to construct a unique blueprint for the future of U.S. culture. Via this blueprint, magazines helped ease the nation's passage to an industrially based society.

Although the current role of magazines in the U.S. might tempt readers to question Schneirov's attention to the medium, the author effectively allays any such doubts. Early on he documents the growth of the Progressive Era's magazine industry (circulation tripled between 1890 and 1905) and the scope of the medium's reach — "65 million or about three magazines for every four people." Indeed, during this period magazines could be considered *the* mass media of the nation. They were an affordable, readable, and widely dispersed source of information and commentary.

What sort of blueprint were magazine editors projecting to their readers? Schneirov argues that dreams of abundance, social control, and social justice comprised the agenda of these editors — three dreams fueled by turn-of-the-century technology's redefinition of activity, motion, and energy as well as by the

changing stance of America's middle class. In essence, America's changing social structure provided a context for magazine editors as they instilled public desire for their own special visions of power and pleasure.

Schneirov's data are impressive. His conclusions stem from an exhaustive analysis of every issue of *Cosmopolitan* magazine published between 1893 and 1914. Schneirov also includes a healthy dose of articles from *McClure's* and *Munsey's* and a more select sample of stories from publications such as the *Atlantic Monthly, Century, Harper's Monthly, Scribner's,* and *Women's Home Companion*. The coverage is broad, giving readers a rich snapshot of the popular philosophies of the era.

In addition to its value as social history, *The Dream of a New Social Order* should be of great interest to sociologists of culture. Its theoretical base is true to Griswold's notions of a "cultural diamond." Schneirov considers the intentionality driving magazine editors (i.e., the cultural producers) and examines the impact of editorial decisions on both the medium's audience and the cultural object itself. Finally, he explores such issues within the special context of the historical times. Schneirov thus provides a valuable application of an inclusive cultural model. Those teaching general seminars in culture may find his book a useful case study illustrating Griswold's theoretical approach. Further, the volume's readability and topic make it suitable for advanced undergraduate or graduate courses in American studies, media history, modernization, and social change.

Politics and Religion in Central and Eastern Europe: Traditions and Transitions. *Edited by William H. Swatos Jr.* Praeger, 1994. 230 pp. $59.95.

Reviewer: MICHELE DILLON, *Yale University*

The essays in this volume provide a useful introduction to the varying ways in which historically embedded interconnections between religion and national identity in several Central and East European countries contributed to and shaped the collapse of Soviet Communism in 1989. Especially welcome is the inclusion in its purview of countries that have received little analytical attention from sociologists and political scientists. Thus, for example, the reader is alerted to the cultural and political importance of religion — whether as oppositional discourse, institutional space for "civil society," or theological belief system — not only in Poland (which is discussed extensively both in this work and in the broader literature) but also in Lithuania, Estonia, Latvia, Ukraine, Georgia, Armenia, Byelorussia, Russia, Hungary, East Germany, Slovakia, the Czech Republic, and the former Yugoslavia.

Irrespective of the different theoretical frameworks employed by the contributors, many of the essays highlight the legitimation problems and resulting practical implications arising from the inevitable interpenetration and mutual adaptation of religious and political institutions. Patrick Michel, in particular, emphasizes what he sees as the "impossibility" of drawing a boundary between the political and religious domains under Soviet Communism. Focusing on Poland, Michel argues that the contradictions inherent in the Soviet system meant that the religious field was doubly solicited by the political regime as a potential instrument of legitimation and by the people as a source of transformation: "the

religious found itself both restrained from entering on an equal footing a field that was not its own and working simultaneously on the consolidation of the regime (the price of institutional survival) and on its destruction, as a vector of disalienation, detotalization, and desovietization (the price of safeguarding its identity)."

As Hank Johnston notes, however, the relationship between the state and the national church varies, depending on historical, geopolitical, and security factors; regime ideology; and control of the local population. He emphasizes that there is no linear relationship between the freedom given to the national church and its affinity with nationalist causes. Church collaboration or complicity with the regime thus has different implications in different cultural contexts — in Armenia, for example, Johnston observes that the church's complicity with the regime did not damage its legitimacy among the population. On the other hand, Natalia Dinello's analysis of Russian public opinion data suggests that "the only institution that inspired considerable trust on the eve of the breakdown of the Soviet Union was the Russian Orthodox Church," a finding she attributes to the fact that "the church was neither involved in the establishment of the Soviet regime nor held responsible for the failures of perestroika."

Similarly, the impact of the collapse of Soviet Communism on religious revitalization varies across different societal contexts, a point stressed by several of the contributors (see especially the chapters by Irena Borowik, Luigi Tomasi, and Ivan Varga). A complicating factor shaping the cultural strength of religious institutions in the postcommunist era is their organizational adaptability to broader intrainstitutional dynamics. Thus Catholic churches in eastern Europe must now confront the challenge of cultural openness articulated 30 years ago by Vatican II (see Varga), while reunited Evangelical German churches must adjust administratively to one another (see Beckley, Chalfant, and Johnson); furthermore, both the Evangelical German Church and the Slovak Evangelical Church (see Gluchman) must carve out a new, morally authoritative identity.

Finally, although religion may well have lost some of its power to enchant, all the essays in this volume serve as reminders that religion continues to provide institutions and social movements with a rich resource and an anchor for culturally contested symbols and meanings. This function is perhaps expressed most concretely by the Bosnian-Croat conflict over "ownership" of Medjugorje, the site of a populist Catholic shrine located in Bosnia-Hercegovina (see Markle and McCrea). This book is a valuable contribution not just to students interested in eastern Europe or religion but to anyone seeking a deeper understanding of the multifaceted ways in which culture both drives and constrains social and political action.

Chinese Families in the Post-Mao Era.
Edited by Deborah Davis and Stevan Harrell. University of California Press, 1993. 370 pp. Cloth, $55.00; paper, $17.00.

Reviewer: YU XIE, *University of Michigan*

This valuable volume includes contributions from five sociologists, five anthropologists, and a psychiatrist. Pooling fresh empirical results from ten independent research projects conducted in different locales in China, the book provides the

latest evidence on recent changes in the contemporary Chinese family and their implications. Written for a broad audience, both China specialists and social scientists with an interest in the family will find it valuable.

Although the essays in this book were first presented at a conference, they are not part of a single, organized research effort. Methods, data sources, and field locations vary greatly. However, Deborah Davis and Stevan Harrell have done a fine job of keeping a coherent focus throughout the volume. Specifically, the book addresses the following question: What are the consequences for the family of the rapid political, social, and demographic changes in China that began in 1978 and continue today in full force? In their introductory essay, Davis and Harrell provide a good historical background and set up a broad conceptual framework for later chapters that recognizes cultural, demographic, economic, and political forces at work and their interplays in affecting processes related to such family outcomes as marriage, childbearing, and caring for dependents.

A review chapter by Jonathan Unger on urban families is followed by ten remaining chapters that are all based on original research on a variety of topics in different settings. There are three chapters on household structure, by Davis (in Shanghai), Harrell (in rural Sichuan), and Graham E. Johnson (in rural Guangdong); three on marriage, by Mark Selden (in rural Hebei), Helen F. Siu (in rural Guangdong), and Martin King Whyte (in Chengdu); two on childbearing, by Susan Greenhalgh (in rural Shaanxi) and Hill Gates (in Taibei and Chengdu); one on caring for family members with schizophrenia, by Michael R. Phillips (in various cities); and one on caring for the elderly, by Charlotte Ikels (in Guangzhou).

In general, the empirical work is well executed and the associated interpretations well thought out. The only problematic essay is the one by Gates on birth limitation among urban capital-owning women. Her thesis is basically Marxian; she argues that among petty-capitalist women the power derived from owning capital positions them to determine their own destiny by limiting their fertility. Her evidence, however, is weak and inconclusive. The data she uses from her interviews with small groups of women do not rule out alternative explanations. For example, economists have long argued that a higher earning capacity increases women's opportunity cost for childbearing; furthermore, the modernization perspective predicts that better-educated women tend to be more in control of their childbearing decisions and that they have lower fertility rates. That is to say, the lower fertility levels that Gates observed among capital-owning women may have had nothing to do with their ownership of the means of production. Gates's tendency to twist evidence to fit her argument is apparent in her statement that "life expectancies for Chinese women were lower than those for men," for which she provides no historical context, leading readers to believe that her remark applies to the post-Mao period. I consulted her source and found that since the 1960s life expectancies for Chinese women have been consistently higher than those for men.

All of the authors base their research on relatively small samples (between 100 and 1,000 cases) that each drew from local communities, which is somewhat surprising as well as puzzling. Why not have made use of existing large-scale national data sources? For example, high-quality unit-record data from the 1982 Chinese census have been available for some time and could have been used by the authors. A few large fertility surveys are also available. While there might be

a subdisciplinary culture opposed to the use of secondary data among China specialists, I see no justification for such opposition. From a statistical point of view, more information should be no worse than less information. In the long run, it is important that China specialists begin to share data and become accustomed to secondary data analyses if they wish to make cumulative progress.

Voicing Social Concern.
By Otto N. Larsen. University Press of America, 1994. 306 pp. $32.50.

Reviewer: KAREN A. CERULO, *Rutgers University*

The traditional publication forum for sociologists holds little opportunity for the placement of essays and commentaries. The professional journals are almost solely restricted to complex theoretical treatises or systematic empirical research. Few homes are provided for pieces that encourage the sociological eye to focus and thoughtfully speculate on the emerging patterns, current events, or long-term trends that constitute our social worlds. Such a glaring gap in the literature makes the appearance of *Voicing Social Concern* especially valuable. Within its pages readers will find 33 considered and often provocative commentaries offered by well-known scholar Otto N. Larsen.

These essays were originally presented as talks during Larsen's worldwide travels. They span a 40-year period and together address a wide variety of topics, including censorship, human relations, innovation and productivity, media violence, political elections, pornography, and the future of sociology. They are straightforward, clear, and void of professional jargon. Indeed, in the author's own words, they represent and attempt to "convert research ideas into a pertinent message of interest to intelligent people beyond those regularly met in a circle of peers with technical specialties."

Those looking for in-depth treatment or complex theoretical excursions will be disappointed by the volume. But those searching for ways to jump-start critical thinking will find it a useful tool. Despite the broad time period from which the writings are drawn, Larsen's essays raise fresh, insightful questions on currently pressing issues. For example, even though his essay on the effectiveness of communication, "Trying to Be Effective — Vying to Be Successful," is over 30 years old, the issues Larson probes remain at the heart of cultural sociology and communication discourse. In some cases the essays exemplify the intellectual creativity generated when one merges nontraditional "data" with key sociological concerns. "Comic Books and Creative Leisure" provides one such illustration. And often these commentaries afford the opportunity to retrace thinking on key issues such as pornography, sex, violence, and sociology itself. In so doing, Larsen's work provides a basis by which to enhance the fruits of future action.

Both the content and style of this book make it suitable for advanced undergraduate courses as well as graduate seminar classes. In each of these arenas the Larsen essays could easily serve as an entree to more in-depth treatments of the matters at hand. And whether one agrees or disagrees with Larsen's views, *Voicing Social Concern* is sufficiently engaging to capture the reader and spark subsequent meaningful discussion.

Interpreting the Field: Accounts of Ethnography.
Edited by Dick Hobbs and Tim May. Oxford University Press, 1993. 247 pp. $24.00.

Reviewer: MARTHA COPP, *East Tennessee State University*

When this book arrived, I envisioned reading some gripping "confessionals." For students and seasoned field-workers alike, fieldwork accounts can be engaging without forcing one to get one's hands dirty or to make sense of the data. They may also fill readers with envy or prod them to improve their work. When confessionals combine scholarship and self-help with an analytical hook, many gobble them up. But as I have learned, when accounts lack analytical purpose, they can bore a reader silly. The eight essays in *Interpreting the Field* offer unequal thrills. Some contributions are excellent and coherent, some contain momentary insights, and others are sedatives.

The authors of this collection represent varied disciplines: sociology, anthropology, criminology, geography. In addition to telling familiar stories of getting in, getting along, and getting finished, they explain the social conditions under which they did their work. All conducted fieldwork in Britain, all tried to be sensitive to politics and social class, and all revealed ethical uneasiness despite covert or overt research practices. The researchers also attempted "reflexivity" by treating themselves as sources of data. Because their essays differ in substance and quality, individual summations follow.

In the lengthy lead essay on studying hooligans, Gary Armstrong narrates reflexively but lacks analytical focus. Dick Hobbs details how the tone and content of his writings shifted to suit different audiences as he gained academic footing. Tim May critiques unemotional traditional ethnography but provides cryptic examples from his work on the probation service. Arguing inconclusively that openly partisan researchers can obtain good data, socialist researcher Penny Green revealed her views only to those she sided with in a miners' strike. With welcome clarity, Clive Norris incisively discusses the ethical dilemmas he resolved while studying the police. Jane Fountain follows with a looser and longer exploration of her admittedly unethical strategies to research cannabis dealers.

Sasha Roseneil's essay on studying the women's peace encampment at Greenham Common provides the book's strongest examples and arguments in favor of reflexive research. And in the poignant final essay, H.L. Ackers demonstrates how she distorted and weakened her analyses by separating her personal life from her research, writing only what an academic audience would deem acceptable. Both Roseneil and Ackers successfully consider how their gender and sexuality mattered (Ackers also considers her race); and, less directly, May looks at how masculinity may have affected his research. As a result, their three essays seem more complete than the other accounts.

Ideally, all the authors should have written reflexive ethnographies. Although some did, many did not, citing the fear and pressure of the rigid academic generalized other. As long as field-workers keep official analyses uncontaminated by "personal" feelings and experiences, they will feel compelled to write separate confessionals, as in this volume. Nevertheless, the essays in *Interpreting the Field* suggest that such practice is becoming increasingly unsatisfactory because the standards of good ethnography now include being reflexive. Recommended for

patient students and established field-workers, the accounts in this collection shed light on how to achieve those higher standards.

Peacekeepers and Their Wives: American Participation in the Multinational Force and Observers.
By David R. Segal and Mady Wechsler Segal. Greenwood Press, 1993. 181 pp. $59.00.

Reviewer: JAMES BURK, *Texas A&M University*

The use of armed forces for peacekeeping has been a matter of increasing interest since the end of the Cold War. Such interest is practical as policy makers try to determine how armed forces may be used to minimize collective violence around the world. This issue is theoretical as well, especially for sociologists. Armed forces are changing institutions. In recent years they have expanded their mission from fighting wars to participating in operations other than war. Yet few careful analyses of how this change has occurred and how it has affected related institutions exist. The new institutionalism has yet to embrace studies of the military. This book is an exception. Working from a social constructivist perspective, the authors have labored since the early 1980s to place the involvement of U.S. forces in peacekeeping into historical context, to indicate the organizational adaptations required to deploy war fighters as peacekeepers, and to trace the consequences of peacekeeping deployments on the wives left behind. (The focus on wives alone is justified, as no women were deployed in the peacekeeping mission studied.)

The book is organized into three parts. In the first Segal and Segal review the history of peacekeeping in the West from the League of Nations to the present and set the stage for understanding the role of the U.S. as peacekeepers deployed in the Sinai between Israel and Egypt. In part 2 they describe an elaborate comparative study of elite paratroopers and light infantry fighters as the latter prepared for deployment, were deployed, and returned from peacekeeping. In part 3 the authors consider how the wives of peacekeepers adjusted to the deployments.

Two major findings follow from this research. First, peacekeeping led soldiers to see the world in more peaceful terms. This was true for both paratroopers and light fighters, although paratroopers retained a more bellicose worldview. Nevertheless, there was no strong relationship between a soldier's combat orientation and his attitude toward peacekeeping. More important to predicting a soldier's acceptance of the peacekeeper's role was the belief that his peacekeeping experience would enhance his career. Second, wives believed the deployment changed their own lives. Most reported increased personal growth and greater independence; and they expected to keep their freedom when their husbands returned home, though many were disappointed. Whether wives adjusted well to peacekeeping deployment depended in part on their coping style, but contextual variables were also important. Those who were well connected to the military community during deployment adjusted better than those who were not. Connections affirmed the importance of the peacekeeping mission, giving meaning and legitimacy to the sacrifices of the well connected, a comfort that the poorly connected lacked.

On balance, the authors of this study have done a good job describing what peacekeeping deployment means for the daily lives of soldiers called on to perform the mission and what it means for their families. It is an essential and welcome contribution that should encourage further study about how the military's use in peacekeeping has altered its institutional self-understanding and its relation to the larger society.

Crime and Its Social Context: Toward an Integrated Theory of Offenders, Victims, and Situations.

By Terance D. Miethe and Robert F. Meier. SUNY Press, 1994. 209 pp. Cloth, $57.50; paper, $18.95.

Reviewer: ROBERT M. O'BRIEN, *University of Oregon*

In *Crime and Its Social Context,* Miethe and Meier attempt to integrate offender-, victim-, and context-based theories of crime into a single theoretical perspective. The criminal act is the dependent variable — not offenders, not victims, and not crime "hot spots" or the social relationships between victims and offenders. To explain the criminal act "requires attention to offenders and to victims, but it can never be complete without a sense of the context in which criminal acts take place." The authors address their task at both the conceptual and empirical levels. Their strategy, much like Cohen and Felson's routine activity approach, pays attention to motivated offenders, suitable targets (victims), and the absence of capable guardians (context) — components often ignored in empirical research.

In the two chapters that follow their introduction, Miethe and Meier describe theories of criminality, victimization, and criminal opportunities. Their summary of these theoretical perspectives is excellent and concise. In chapter 4 they integrate and enhance the perspectives. The model of criminal events that emerges includes sources of offender motivation, victim characteristics that provide criminal opportunities, and the social context (physical locations, interpersonal relationships, and behavioral settings) as factors affecting the probability of criminal events. The authors emphasize that these factors interact to create the probability of a criminal act.

In chapters 5 through 8 Miethe and Meier examine data from a variety of sources: uniform crime reports, national crime surveys, the U.S. census, and their own victimization study. Sometimes they make simple descriptive comparisons of results from these varied sources; at other times they make overtime comparisons; and at still other times they compare census tract–level analyses with city-level and SMSA-level results. Space limitations preclude reporting these results, which come from different analytic techniques, but we should be encouraged by similarities in the findings at several levels.

In chapter 9 the authors exploit their data from a Seattle telephone survey. This survey is unique in its range of questions and in its sampling design, which allows estimates of block-level characteristics to be computed. The authors first present logistic regression analyses of an individual's risk of victimization for various types of crimes. These equations include several individual-level variables, such as age, gender, race, nights spent out, home unoccupied, dangerous activities, living alone, and family income. They then add several

block-level characteristics, such as residential mobility, busy places, ethnic heterogeneity, low socioeconomic status, mean nights out, and mean value of carried valuables. The authors find that both individual-level variables and contextual (block-level) variables are important determinants of an individual's risk of victimization. For example, a person's risk of being robbed or mugged is negatively related to that person's family income and positively related to the socioeconomic status of the block on which he or she lives That is, family income is associated with less risk of an individual falling victim to these crimes, even when controlling for the socioeconomic status of the block on which the person lives. For burglary and auto theft, both the individual's family income and the socioeconomic status of the block on which the person lives are related positively to victimization risk. These are indeed interesting and meaningful findings.

Next the authors examine interactions by dichotomizing context-specific block-level variables, such as family income (low/high) or busy places (low/high). They find that "having safety precautions" lowers risks of burglary in high-income but not in low-income areas. Safety precautions do not lower burglary risks in busy places, but they do in less busy ones. These and several other interactions demonstrate the importance of context in determining which factors influence the risks of crime.

Overall this is an excellent book. Miethe and Meier demonstrate that many empirical analyses of rates of crime have been misspecified, because the analyses have not taken into consideration *simultaneously* criminal motivation, victim availability, and the social and physical context in which criminal acts occur.

A Festival of Violence: An Analysis of Southern Lynchings, 1882-1930.
By Stewart E. Tolnay and E.M. Beck. University of Illinois Press, 1995. 297 pp. Cloth, $49.95; paper, $19.95.

Reviewer: KENT REDDING, *Indiana University*

Between 1880 and 1930 mobs lynched between 3,000 and 4,000 southerners, most of them African Americans. Since the end of the lynching era dozens of books and scores of articles have been written about one or more of those mob murders. Although Tolnay and Beck's important book is unlikely to stabilize what appears to be a burgeoning scholarly interest in this grisly phenomenon, it is by far the most sophisticated and sustained quantitative analysis to date of the structural correlates of southern lynchings.

Readers who have followed the authors' spate of articles on the topic in various journals will not be surprised by many of the findings reported in their recent book. Still, *A Festival of Violence* is no mere compilation of previously published work, nor is it a dry reportage of statistical findings. While the writing is sometimes plodding and repetitive, the authors generally do an effective job of integrating the various strands of their previous work and of dramatizing their concerns by weaving into the text narratives of particular lynchings.

Their basic concern is explaining the considerable variation in lynchings across 50 years and 800 or so counties in 10 states (Texas and Virginia are omitted because of data problems; Kentucky is included as part of the South). Initial chapters provide historical background and descriptive statistics and sketch the

basic conceptual model. The heart of the book lies in four chapters, three devoted to testing components to the "threat perspective" of race relations and one to explaining the demise of lynching by 1930.

Using a variety of specifications of time-series and cross-sectional models of lynching, the authors reject both popular-justice and political-competition explanations for lynchings. Whereas contemporaries claimed that lynchings were a form of popular justice, Tolnay and Beck find neither a reinforcement nor a substitution effect of lynchings for executions in either time-series or cross-sectional analyses. Moreover, they report a *negative* association between Republican (and even Populist) strength and the number of lynchings, and no significant change in lynching patterns after the implementation of formal disfranchisement measures.

Economic-threat models, derived from Blalock, turn out to provide the best explanations of African American lynching patterns. Tolnay and Beck find lynching associated with lower cotton prices over time, cotton dominance and white landlessness across counties, and seasonal demand for labor. These associations reflect both the need of white planters to control black laborers and the economic competition between poor whites and blacks.

Finally, Tolnay and Beck argue that the number of lynchings declined not because of antilynching activism, the advent of New Deal agricultural policies, or improvement in law enforcement. Rather, they find a reciprocal relationship between the great out-migration of African Americans and lynching in the 1910s and 1920s: lynching increased out-migration, and the increase in out-migration resulted in fewer lynchings, as fear of labor shortages may have induced white planters to curtail mob violence.

Fortunately, in the authors' hands conventional statistical techniques are not used as blunt instruments but as tools for testing and contextualizing theoretical and historical explanations. By using their various statistical models in historically sensitive ways, they have significantly advanced our understanding of the patterns of lynching without ignoring the complexities of southern history.

That said, historian W. Fitzhugh Brundage's recent work on lynchings in Virginia and Georgia may lead one to wonder if Tolnay and Beck are justified in treating all lynchings the same, given the different social organization of the mobs used to accomplish them. Contrary to what one might have surmised from having read Tolnay and Beck's book, according to Brundage mass mobs with broad community support were involved in a minority of lynchings in Virginia and Georgia.

Furthermore, in spite of Tolnay and Beck's skillful and nuanced use of regression models, those used in historical research have limitations. In this case they explain the correlates of lynching without explaining the processes, either of how lynchings became the established means for controlling the economic threat of blacks to whites, or of how they were sequentially related to one another. But such problems await new historical research and the application of new methods. In the meantime, we have *A Festival of Violence* as a fine and definitive work on the social-structural underpinnings of the savagery of southern lynchings.

The Anti-Abortion Movement and the Rise of the Religious Right: From Polite to Fiery Protest.

By Dallas A. Blanchard. Twayne Publishers, 1994. 177 pp. Cloth, $26.95; paper, $14.95.

Reviewer: CLYDE WILCOX, *Georgetown University*

Dallas Blanchard opens this book with a quotation from the *Reader's Digest* that makes it clear that he believes that the pro-life movement is dangerous and that he has an obligation to warn his reader of that danger. It also exemplifies to an extent the author's approach to documentation: a few hours of research would have uncovered a number of quotations from more important sources that would have made his point more clearly. A serious scholarly study of the rise of this diverse, complex social movement would be welcome, but this book appears to have been written hastily, and Blanchard's personal bias frequently comes through.

An example of hasty production can be found in the author's presentation of the history of the burning and bombing of abortion clinics. Blanchard first reveals data indicating that 1992 is *the* crucial year in determining a possible trend toward increasing violence, but he later neglects to include data from 1992 in tables in which he outlines information on arson and bombings separately, although the data are from the same source.

An example of Blanchard's bias occurs in chapter 4, where he writes: "Fundamentalism, then, worships a violent God. . . . The fundamentalist mindset espouses physical punishment of children, the death penalty, and the use of nuclear weapons; fundamentalists are more frequently wife abusers, committers of incest, and child abusers." Although fundamentalists do endorse spanking, capital punishment, and maintaining a nuclear arsenal as a deterrent, the same can be said for a majority of nonfundamentalist Christians, who presumably worship a more peaceful God. As documentation for his claim that fundamentalists abuse their wives and children, Blanchard references an unpublished 1988 conference paper and a 1960s psychiatric study of some 60 families of widely diverse religious backgrounds in which the authors claim the following: "*It is our impression* that *among those who were actively involved in their religion*, there was a greater than average adherence to . . . fundamentalist religion" (emphasis added). This is sparse support for a bold assertion.

Blanchard frequently blurs important distinctions — between the potential constituency of the pro-life movement and actual organizations and activists, between the Christian Right and the pro-life movement, and less frequently between Catholics and fundamentalists in the pro-life movement. For example, he argues that "the earliest explanation for the movement was that participants were members of the working class attempting to shore up, or defend, their declining social status." As support for his argument he cites Gusfield's classic book on the temperance movement and Lorentzen's analysis of the school book controversy in Charleston, West Virginia, but neglects research in which this explanation in used in relation to antiabortion groups.

When Blanchard attempts to get at the motivations of the actors in the movement, his argument is confusing. His list of "fundamentalist issues" is not a list of fundamentalist positions on reproductive issues, for it includes foreign

policy, but it is also surely not a list of major issues for fundamentalists, since it excludes education and is heavily weighted to reproductive issues, including contraception, which Blanchard incorrectly claims fundamentalists oppose. What precisely is this list supposed to represent?

Blanchard provides an overview of the major social movement theories that could explain the pro-life movement, as well as a history of abortion and the rise of the pro-life movement. But his coverage is superficial and leaves interesting questions unanswered. For example, soon after stating that the initial reaction to the *Roe* v. *Wade* decision in 1973 came from Catholics, he writes that most opponents of the decision are Protestant fundamentalists. How did orthodox Protestants become involved in what was initially perceived by many to be a "Catholic issue"? How did pro-life organizations grow and proliferate?

The most valuable part of the book is a list of major pro-life organizations and their addresses in the appendix. Blanchard also provides some descriptions of the organizations in the text, although they are uneven and often drawn from secondary sources.

Inventing the Feeble Mind: A History of Mental Retardation in the United States. *By James W. Trent*. University of California Press, 1994. 356 pp. $30.00.

Reviewer: BRUCE BELLINGHAM, *Florida State University*

Clearly the authoritative work on the institutional history of mental retardation from the beginnings of special education to the present, *Inventing the Feeble Mind* is also a fine example of "cultural constructionist" policy analysis. Without glib nominalism, James W. Trent treats retardation as a social fact, and his analytic treatment is enlivened by a mature sensitivity to moral issues. Teachers of social problems, health policy, or historical methods who expose advanced students to the constructionist research tradition could hardly find a better exemplar.

Eighteenth-century idiots enjoyed a protected place in local communities, but by the Jacksonian era, Trent reports, vocational complexity and urbanization made mental impairment more awkward. Yet modernization also brought the promise of amelioration. Scientific educators like the expatriate French "physiological" pedagogue Edward Sequin — whose experiments and methods still constitute the greater part of special education's repertoire — made brilliant progress. These charismatic Victorian special educators, working in intimate contact with small groups of resident pupils, made miraculous progress with apparently hopeless cases. Nonetheless, their successes commenced more than a century of custodial warehousing and gross abuses of power.

Different epochs in the prolonged tragedy of retardation are marked by a sequence of governing images or rhetorics. Trent links these shifting images to the evolving professional strategies and ambitions of the medical directors claiming jurisdiction over idiocy. Economic elites also played a surprisingly active role in shaping policy and the human sciences of impairment. Links between the practical and ideological enterprise of mental hygiene and the culture of the natural sciences pose sociology-of-knowledge issues that Trent handles with aplomb.

A summary simplifies Trent's subtle arguments. Briefly, a vast program of institution building was warranted by the useless idiot who burdened the productive capacity of the natal family and community in the rebuilding effort after the Civil War. The criminal (male) or sexually profligate (female) hereditary "menaces" of the Progressive Era eugenics scare implied broad social control of immigrant and lower-class populations. The normalizable idiot of the interwar years was a creature of the need to cycle "paroled" inmates out of institutions in response to higher intake rates and of a professional desire for new roles in community mental hygiene. In 1970, wrongly incarcerated mentally challenged citizens were discovered coincident with prohibitions on using inmates for unpaid in-house labor and with fiscally motivated state wishes to release them — thereby off-loading special-education costs to the federal government, which has sustained the custodial bias through payments to private service providers. No coherent attempt was ever made to define or serve the interests of the retarded themselves.

Paradoxically, the germ of this failure traces beyond the bustling "medical man" managers of the late nineteenth century to the heroic Victorian special educators. The Victorians did some good, but by adopting a limited medical model of impairment and by focusing on particular symptoms of defect and discrete therapeutic techniques, they defined the idiot and his or her social and economic roles and prospects as therapeutically irrelevant. A limited, technical vision facilitated exaggerated professional claims, consequent institution building, and routinization of training. It also resulted in substantive irrationality being inserted into policy: apart from smooth institutional functioning, mental hygienists could not state the purpose of treatment. The question of purpose raised broad social issues about family life and labor markets that medical officers kept at bay for over a century.

That such a faulty project endured suggests that the asylums offered solutions to practical problems arising in the community and family. Trent skirts demand-side agency, however, stressing instead medical mystique and salesmanship. One poignant but untheorized discussion of parental estrangement from institutionalized children offers a tantalizing glimpse of the richness of the records (as do the eugenists' own distorted "case histories"). But questions from the interactionist tradition of sociology are missing. For example, at midcentury idiots were all males. Hereditarian discourses of sexual dangerousness later warranted recognition of female idiots, Trent argues. Interactionist accounts of the micropolitics of trouble preceding deviant classifications suggest that the changing character of domestic conflicts involving girls might have had more to do with the invention of female idiots than scientific models did. Indeed, intake pressures are noted as a prod to entrepreneurial innovation, but the sources of this pressure remain mysterious. Trent cannot explain why female idiots were suddenly recognized because situated interactions connecting family and community to the institution are beyond the boundaries that constructionism imposes on itself. A history of the retarded has still to be written. But this history of their keepers is splendid.

Paths to Homelessness: Extreme Poverty and the Urban Housing Crisis.
By Doug A. Timmer, D. Stanley Eitzen, and Kathryn D. Talley. Westview Press, 1994. 210 pp. Cloth, $55.00; paper, $16.95.

Reviewer: JAMES WRIGHT, *Tulane University*

Does homelessness stem from the personal deficiencies of homeless people or from problems and dislocations in the larger social structure? Some two dozen volumes published in the last decade argue in favor of the latter viewpoint, which most sociologists would find self-evident. Given the quantity of recent material on the homeless, one wonders whether anything worthwhile remains to be said.

The value of *Paths to Homelessness* is not that it contains major new insights; social scientists conversant with recent literature on homelessness will find themselves in familiar terrain. Timmer, Eitzen, and Talley argue that homelessness results from an increasing rate and depth of urban poverty coupled with a precipitous decline in the supply of low-cost housing; if anything, federal policies on housing, income, welfare, and homelessness have aggravated the problem. Explanations that focus on the personal disabilities of homeless people — their rates of alcohol and drug abuse, mental illness, social estrangement, and the like — are rejected as "blaming the victim." The authors thus present a very different line of argument from that advanced by Baum and Burnes in their 1993 book, *A Nation in Denial* (which, ironically, was also published by Westview Press though not cited in the present work).

What distinguishes this volume from its many worthy competitors is the authors' coupling of social-structural insight and argumentation with qualitative data derived from ethnographic interviews with homeless people in Tampa, Denver, and Chicago. Unlike studies based on "impersonal" survey data, this one gives voice and humanity to homelessness by letting homeless people speak on their own terms. Unfortunately, the authors command neither the ethnographic brilliance of Elliot Liebow (*Tell Them Who I Am*, 1993) nor the journalistic rage of Jonathon Kozol (*Rachel and Her Children*, 1988), both of whose evocative portrayals succeed as literature and as social science. In comparison, *Paths to Homelessness* is rather flat and ultimately unsatisfying.

One concern is that the authors' efforts to create sympathy for the homeless will have an opposite effect on many readers. For example, in a discussion of "middle class moralisms about appropriate behavior" that shelter providers frequently attempt to enforce, the apparently incredulous authors tell the readers that "one of the shelters instituted a policy of not paying for children's drug prescriptions if their parents were spending money on cigarettes." In the same vein, the authors relate the following: "*Interviewer:* How do you feel about the way the shelter treats you? *Diane:* I mean its not like livin' in luxury. You have to like almost earn your keep if you stay here." Diane's reference is to the obligatory chores that people who live in the shelter are expected to perform: make their own beds, wash dishes, mop floors. She adds a bit begrudgingly, "The dishes have to be done by somebody, so it might as well be us because we ate." I have no doubt that the authors meant to depict Diane sympathetically, but in their portrayal she comes off as an ingrate.

Diane is not an isolated example. Many of the homeless people whose stories are recounted here appear less as the unfortunate victims of large-scale social

forces than as foul-mouthed whiners who expect simply to be given whatever they think they need. In more than a few cases, the authors' attempts to dismantle stereotypes inadvertently reinforce and amplify them, a problem aggravated by what might be called the "shit-fuck-asshole-cocksucker" approach to urban ethnography, where the informants' every vulgarity is faithfully reproduced — a style more tedious than offensive.

Paths to Homelessness is a capable overview of the arguments and research literature depicting homelessness as a social-structural problem. Readers seeking the voice of the homeless, however, will be better served elsewhere.

World Changes in Divorce Patterns.
By William J. Goode. Yale University Press, 1993. 354 pp. $35.00.

Reviewer: FRANCES GOLDSCHEIDER, *Brown University*

Goode has given us a treasure — a systematic description of trends in various aspects of divorce for many of the world's regions between 1950 and 1990. He provides a wide array of information, drawn mostly from official statistics and documents. His book contains an introduction, a conclusion, and nine chapters that each address a particular grouping of countries (with groupings based primarily on similarity of institutions, culture, and history). Most countries for which reasonable data on divorce over time exist are included in the study.

Europe gets four chapters: the Nordic countries, the Catholic south, the (formerly) Communist East, and a core residual, which is "the general pattern" for the world. These chapters are followed by two on Europe's great areas of cultural conquest: the English and Iberian blocks. The rest of the world is represented in three chapters containing special groupings: Japan and Malaysia, the "Arab world," and Indonesia, Taiwan, and China.

The topics covered in each chapter necessarily vary to a certain extent, but Goode generally includes information on changes in divorce rates, the legal system, the average duration of unions ending in divorce, children's involvement, remarriage and cohabitation, and postdivorce arrangements. This framework works well for the most part, allowing Goode to be both thorough and particular.

So what is not in the book? That is not a fair question, since every piece of research is necessarily selective. Given that "world" is included in the volume's title, however, it is reasonable to raise questions about the author's choice of countries. Further, on behalf of the generations of family scholars who learned about "convergence" from Goode, it is important to address the theoretical framework of his study.

Goode tells us that the data do not allow comparable analysis (even at the level of the Latin American chapter) for Africa or southern Asia. The same probably holds for the adjacent countries, including much of Southeast Asia. Yet Thailand has had a modern data-collection system throughout most of the period reviewed and may represent its more disrupted neighbors on many dimensions, but it is not mentioned. More puzzling is the omission of Greece. Presumably these countries did not fit the grouping framework.

However, William Parrish has provided a reconsideration of Goode's framework, which suggests that the more countries included the better. Using basic

regression analysis, Parrish shows that Goode's nine groupings really come down to three sets clustering around very tight regression lines. One group has responded little or slightly negatively to increases in conventional indicators of industrialization (the Arab and Moslem countries), and two have responded positively to them (Eastern Europe, starting from a high level, and the rest of the world, starting from a lower level).

What accounts for these patterns? Why should divorce rates decline in some places before rising, which Goode correctly indicates was even the pattern in Europe under Catholicism? Perhaps if Goode had thought as systematically about changes in men's familial roles in the earlier period as he did about those in women's in the modern period, he might have made more connections. His older analysis of the convergence on the nuclear family may provide a key. Perhaps the growth of the nuclear family increased the costs of divorce for women and children. Furthermore, as early industrialization made men increasingly less dependent on family ties and family labor, other still powerful institutions such as the church increased the stringency of divorce in order to tie men effectively to their now more vulnerable wives and children. As women have gained a similar independence with continued industrialization, however, they have developed a greater acceptance of divorce — but as Goode points out, we still have to worry about the children.

Women in Ministry: Receptivity and Resistance.
By Edward C. Lehman Jr. Joint Board of Christian Education (Melbourne, Australia), 1994. 156 pp. Paper, $11.50.

Reviewer: KEVIN J. CHRISTIANO, *University of Notre Dame*

The most satisfying aspect of this small paperback volume about the introduction of female clergy into two Australian denominations is that its author says exactly what he means and clearly means what he says. With this book Edward C. Lehman Jr., an experienced sociologist of religion who has studied the receptions accorded women ministers in the U.S. and Great Britain, extends his research Down Under. Throughout this work he is precise in claiming what he has accomplished at the same time that he is unpretentious in discussing its implications.

"The goals of the undertaking," as Lehman lists them without fanfare in the preface, "were to describe patterns of receptivity and resistance to women as ordained clergy and to identify factors that could explain those differences in attitude" among parishioners in five of Australia's largest states. Of course the obstacles facing women in traditionally male-dominated professions have received extensive coverage in Australia's popular media, as they have elsewhere. And as elsewhere, much of the concern of church leaders in Australia has centered on how — or even whether — age-old disciplines might be modified from the top to better accommodate the aspirations of female members. Scholars of religion have given less attention, however, to measuring the reactions of average churchgoers to the integration of ministries by gender. Lehman's intention, then, was to produce "a study of what ordinary lay church members thought and felt about the issue" of women in the role of priest.

For his national survey, Lehman chose the two largest Protestant denominations in Australia, the Anglican Church and the Uniting Church. The former, as its name implies, derives from the Church of England; the latter is the outcome of a merger in 1977 of Congregational, Methodist, and Presbyterian churches. Initially the two denominations differed markedly in their official postures toward female clergy: whereas the Uniting Church made acceptance of ordination for women an explicit "basis of union" from the outset, the Anglicans, after great controversy and internal division, only recently voted to permit individual dioceses to open the priesthood to women. They still have no policy that is nationally binding.

In his presentation, Lehman moves quickly from simple descriptions of the data to an exposition of bivariate relationships, and from there to statistics generated by multiple regression. Yet he remembers that many of his readers are not social scientists, and so he strains to frame his interpretations in cautious and nontechnical language.

The varied and complex nature of Lehman's results renders an easy summary impossible. However, Australians "as a whole" are "accepting" of women clergy. A minority prefer men to perform priestly functions, but most church members do not care about gender. Instead, they evaluate ministers without obvious reference to sexual stereotypes. "These patterns bode well both for clergywomen and the churches," observes Lehman. "Lay church members show little inclination to burden women ministers with jaundiced perceptions of their ability to do the job."

Lehman finds that opposition to the ordination of women among lay members in the Protestant churches of Australia is connected to conventional attitudes toward sexuality, the negative opinion of a member's own pastor on the subject, and a low degree of salience for the issue itself. More specifically, "it is difficult to overestimate the influence priests and pastors can have on their lay church members." "What they say about the ordination and placement of women as priests and pastors tends to carry more weight than any other single factor."

This book is not a full-length monograph on perceptions of the clerical profession in Australia. Rather, most of the analysis reads like the final memorandum of a researcher to his or her client. Alert readers nevertheless will learn something in passing about the structure of Australian religion. In addition, Lehman furnishes a valuable, if short, section of comparisons between the results of his Australian survey and those of his research in other societies.

Finally, thanks to the care with which Lehman recounts his research process and to appendixes in which his original questionnaires (with marginal frequencies added) are reprinted, readers who have no direct interest in gender, occupations, or religion will at least gain a brief tutorial on the conduct of a small-scale survey project.

Leaving Home Before Marriage: Ethnicity, Familism, and Generational Relationships.
By Frances K. Goldscheider and Calvin Goldscheider. University of Wisconsin Press, 1993. 242 pp. $27.50.

Reviewer: DANIEL T. LICHTER, *Pennsylvania State University*

For most of American family history, young adults have typically resided with their parents until marriage. In *Leaving Home Before Marriage,* Goldscheider and Goldscheider argue that independent, nonfamily living is now a normative though fragile stage of young adulthood. Basing their findings on the 1980 High School and Beyond Survey, they report that nearly 70 percent of high school seniors expected to live apart from their parents before marriage. Indeed, about 50 percent actually became residentially autonomous within six years of graduating from high school. Another 30 percent married, and 20 percent remained residentially dependent on their parents.

This book has several interesting features. It is probably the first comprehensive treatment of premarital residential independence, arguably the best marker today of the transition to adulthood. Goldscheider and Goldscheider dispel the popular view that young adults are returning to the nest in large numbers; unmarried young adults today are less likely than those in the past to live with their parents. The authors examine the important influences of parents' expectations and financial resources on independent living among their children. Finally, they pit economic explanations that emphasize the effect of affluence on residential autonomy against cultural explanations that stress ethnicity, familism, and egalitarianism.

Although *Leaving Home Before Marriage* is useful, well written, and provocative, the analyses and results sometimes provide a weak basis for strong conclusions. The main one is the most overdrawn and contentious. The authors give only modest support to the hypothesis that rising affluence has allowed young adults to "buy" independent living and privacy. Instead, they conclude "that affluence is much less important than values" and "that nonfamily living has simply not been as well accepted in more familistic ethnic groups and more familistic religious denominations as it has been in the dominant, largely secular culture." However, the evidence they present to support the emphasis on values over affluence is weak.

First, the authors measure parental "affluence" rather than the changing economic circumstances of the young adults themselves (i.e., employment, earnings, and school enrollment). Does going to college, finding a job, or becoming unemployed affect premarital residential independence? Do earnings or access to welfare during young adulthood? These are obvious questions to answer before asserting that values predominate over economic circumstances in influencing nest-leaving.

Second, my interpretation of the results is that parental affluence (as measured by socioeconomic status) has a large — not small — effect on nonfamily living. Although parental socioeconomic status is only minimally related to parents' expectations about their children's nest-leaving, it is strongly related to actual behavior. Among children in the lowest decile of socioeconomic status, 50 percent achieved independence within six years of high school graduation,

compared with 81 percent among those in the highest decile. In the bivariate analysis, parental socioeconomic status had the largest absolute effect on premarital independent living — larger than the effects associated with any of the measures for values (i.e., familism, secularization, or egalitarianism).

Third, rather than measuring values directly, the authors use demographic variables as proxies: familism is measured by ethnicity and region; secularization is indicated by one's religious affiliation and denomination; and egalitarianism is measured by gender. Truly convincing conclusions about the importance of values over economics requires a thorough analysis of attitudes about the desirability of marriage; the need for personal privacy, autonomy, and fulfillment; past parent-child relationships (i.e., conflict); and gender roles (i.e., attitudes about female dependence). Indeed, the multivariate analyses indicate that only about 10 percent of the variance in nonfamily living can be explained by the proxies. More and better measures of affluence and values are needed to support the kinds of strong arguments Goldscheider and Goldscheider make.

Fourth, the analysis is based largely on seniors interviewed in the 1980 High School and Beyond Survey; dropouts are excluded. The problems of bias associated the exclusion are potentially serious, in large part because dropout rates differ significantly across various racial and ethnic groups. In 1980, for example, about one-third of Latino males (aged 16 to 24) were high school dropouts, compared with about 15 percent of white males. Can the authors then claim that familistic ethnic groups — like Latinos — are less likely than other ethnic groups to become residentially independent before marriage? How do living arrangements change when Latino men drop out of school? Do they continue to live with their parents? Do they drop out of school to get married or to work? And how does their reason for dropping out affect routes to residential independence? Dropouts cannot be ignored in a self-proclaimed "comprehensive" treatment of premarital nonfamily living.

Despite its limitations, *Leaving Home Before Marriage* is interesting and thought-provoking. The authors draw attention to an increasingly significant life-cycle stage of young adulthood and provide a point of departure for additional research on many relevant questions about the timing, causes, and economic implications of nest-leaving.

Feminist Organizations: Harvest of the New Women's Movement.
Edited by Myra Marx Ferree and Patricia Yancey Martin. Temple University Press, 1995. 474 pp. Cloth, $49.95; paper, $22.95.

Reviewer: REBECCA E. KLATCH, *University of California–San Diego*

This is a collection of 25 essays originally presented at a 1992 conference on feminist organizations. Case studies are wide-ranging and include examination of the Catholic Church, the Chicago Women's Liberation Union, domestic violence organizations, feminist health clinics, rape crisis centers, post–World War II black women's groups, and antifeminist organizations.

The contributors raise important questions about such fundamental issues as the possibilities and dilemmas faced by feminist organizations interacting with the state, the prospects of organizing across class and race, and the inherent

conflicts involved in building coalitions. Many call into question the dichotomy of radical versus mainstream, challenging the assumption that becoming an insider necessarily means co-optation.

Some essays are more theoretically provocative than others. Reinelt's case study of the battered women's movement in Texas raises key issues regarding the tension between the autonomy of feminist organizations and engagement with mainstream institutions. Staggenborg's essay is exemplary in her linking of empirical cases with theory. Analyzing what types of feminist groups lead to what kinds of outcomes, Staggenborg persuasively argues that the notion of organizational success needs to be expanded.

Other chapters are interesting as empirical case studies: Farrell offers a fascinating examination of the tensions faced by *Ms* magazine in its attempt to balance the feminist interests of readers with the interests of capitalist advertisers. Whittier's study of microcohorts in the Columbus women's movement illustrates how turnover of membership redefines feminist collective identity.

One of the most interesting sections of the volume focuses on emotions within organizational life. In an illuminating essay, Verta Taylor makes a powerful argument for recognizing the role of emotions along with instrumental bases for feminist protest. She mentions but does not discuss the negative side of the emotional life of organizations — for instance, how demands for ideological purity can turn women against one another. Luckily, Morgen's examination of feminist health clinics does address the issue of infighting and conflict. In fact, to its credit, the book as a whole is not based on the assumption that if only women led organizations the world would be different. Several case studies demonstrate how women, too, can dominate or reproduce hierarchy.

I have only two complaints. First, more of an effort could have been made in several chapters to draw out the implications of the case studies for organizational and social movement theory. Clearly some essays are better than others at accomplishing this goal. Cheryl Hyde, for instance, does a commendable job of analyzing the varying strategies adopted by nine feminist social movement organizations in countering the antifeminist New Right. The volume also would have benefited from more essays like those by Gelb and Acker, who synthesize main themes and outline directions for future research.

My other problem with the book is that its authors never really come to terms with one of the main questions: What qualifies as a feminist organization? Does a women's bank count as feminist even though the women involved in it do not identify with one another and know little about feminism (see Tom)? Or is it style that makes an organization feminist? Barnett argues that an emphasis on participatory democracy, community, mutual respect, and self-transformation qualifies a group as feminist. But what about explicit goals aimed at changing women's place in the social structure? While it is essential to recognize the importance of class and race, is not some recognition of the role of gender essential for an organization to be feminist? Based on her study of Mexican-American community activists, Pardo argues that groups of women whose members simply become public actors and extend the mother role to incorporate community issues are implicitly feminist, even though many of these women accept the traditional division of labor and assume that men are natural leaders. By this definition right-wing organizations might also be considered feminist. As Jo Freeman contends in the conclusion, empowering women does not alone

constitute feminism. On the other hand, Christensen-Ruffman argues that women engaged with mainstream political institutions should not be considered part of "women's political culture." Such a narrowing of the term merely adds constraint rather than complexity to our understanding of women's politics.

Yet even these problems are provocative in raising critical questions about the meaning of feminism and how to achieve social change. In short, there is much to be gleaned from these instructive essays not only for academics interested in gender, social movements, organizations, and the state but also for activists and practitioners who sustain feminist organizations.

The Selling of Contraception: The Dalkon Shield Case, Sexuality, and Women's Autonomy.

By Nicole J. Grant. Ohio State University Press, 1992. 240 pp. Cloth, $29.95; paper, $15.95.

Reviewer: JULIE BRINES, *University of Washington*

Struggles over women's sexual and procreative rights often take the form of struggles over contraception. As Nicole Grant documents in *The Selling of Contraception*, the story of the Dalkon Shield intrauterine device is an account of one such struggle and its life-and-death consequences. Heralded in the early 1970s as the "superior modern contraceptive," the Dalkon Shield was eventually linked to at least 17 deaths and countless cases of miscarriage, infertility, and internal injury. The case has had a chilling effect on the domestic manufacture and distribution of IUDs. Today only a few such devices are available in the United States; meanwhile, international family planners continue to advocate their use.

In her compelling account, Grant first weaves a social history of modern contraception through the particulars of the Dalkon Shield case (chapter 1 and 2). Representatives of various interests — professional, state, corporate, cross-national — populate the pages of this history, and there is more than enough evidence of greed, ignorance, and malignant neglect to go around. Grant then turns to the testimonials of a "theoretical sample" of 17 women, most of whom used either the Dalkon Shield or another IUD during the 1970s (chapters 3 and 4). Nearly all of these women seemed aware that the use of an IUD involved some risk. Nonetheless, nearly all actively sought out this method of birth control.

Why did these women opt for IUDs over other forms of contraception? Oral contraceptives were rejected at the time because of their links to cancer, stroke, and side effects. Most of the women, though, still desired a birth control method that separated the contraceptive act from coitus. "Coitally independent" methods of contraception, particularly the IUD, relegated fertility control to backstage work, promising to clear the front stage for "fully spontaneous" sexual expression. In short, Grant finds that for the women in her study the decision to use an IUD was linked to a search for sexual autonomy, albeit one bounded by norms that define fully realized heterosexuality in terms of coitus.

Although Grant's analysis is intriguing, her general conclusions are best regarded as provisional (chapter 6). Consider her two primary research questions: "Why did women of diverse backgrounds use the Dalkon Shield?" and "Was this

case an aberration, or did it reflect normal relationships and routine practices within its sociohistorical context?" Grant's sample cannot address these questions. Most of her participants were recruited through a snowball procedure, the mechanics of which tend to attenuate precisely the kind of diversity (variation) among sample members that Grant needs to address her first question. Moreover, evidence indicates that the experience of an IUD-related injury motivated most of the women to participate in Grant's study (75 percent of IUD users in the sample were injured), introducing a selection bias that surely affects whether the testimonials reflect "normal relationships and routine practices." Also troubling is the inclusion of only one woman in the sample who had never used an IUD. The absence of variation makes it impossible to determine whether, for example, IUD users had relationships with their physicians or concerns about sexuality that departed from or resembled those of other women.

Nonetheless, the study raises important issues. As Grant argues in her conclusion, sex will never be safe for women as long as it is viewed as requiring tradeoffs between competing risks (of pregnancy, disease, or birth-control injury). Such tradeoffs, however, seem necessary for heterosexually active, fertile women only in cultures in which sexual activity is centered on the practice of vaginal-penile intercourse. Consideration of alternative heterosexual practices might encourage these women to think of the risks of pregnancy or contraceptive injury in absolute rather than relative terms. It might also remind us of how powerfully norms can frame "choice" or "autonomy," even in the ostensibly private domains of sexuality and reproduction.

Prostitution: An International Handbook on Trends, Problems, and Policies.
Edited by Nanette J. Davis. Greenwood Press, 1993. 403 pp. $79.50.

Reviewer: VALERIE JENNESS, *Washington State University*

The sheer number of books on the history and social organization of prostitution sitting in any university library attests to an ongoing preoccupation with the subject, one that the social sciences and humanities have enjoyed for almost as long as each has been in business. Given this preoccupation, and having contributed to it myself, when I received a 400-plus-page handbook on prostitution to review, I found myself thinking, Does the academy, much less the world, really need yet another book on prostitution? A review of *Prostitution: An International Handbook on Trends, Problems, and Polices* ensures that the answer is a resounding yes.

Nanette J. Davis, a premier scholar in the study of prostitution in the United States, has edited nothing short of an encyclopedic volume focusing on what Gayle Rubin, in an essay published in *Towards an Anthropology of Women* (1975), has identified as the lowest order of the sex/gender system: female prostitutes. As in Vern Bullough and Bonnie Bullough's classic work *Prostitution: An Illustrated Social History* (1978), a wide range of societies that differ in their cultural, religious, social, and legal systems are represented. Written by different scholars from countries across the globe (except Africa and the Middle East), specific chapters detail the social organization of prostitution in Australia, Brazil, Canada, China, England and Wales, Germany, Italy, Japan, the Netherlands,

Norway, Portugal, Singapore, Taiwan, the United States, Vietnam, and (what used to be) Yugoslavia. Each author covers — to greater or lesser degree — nine interrelated themes: social and legal definitions of prostitution, the history and trends of prostitution, the social organization of prostitution, theories of prostitution, the contemporary status and lifestyles of prostitutes, law enforcement related to prostitution, contemporary politics surrounding prostitution, enacted and proposed interventions into the world of prostitution, and policy provisions related to prostitution.

In covering vast geographical and substantive territory, the book "makes no claims for universal representation about prostitution." Nonetheless, anyone reading it will be exposed to both interdisciplinary and international thinking and theorizing about prostitution; a plethora of empirical facts about the history and social organization of prostitution in countries (and cultures) that are very similar as well as in those that are unimaginably dissimilar; the range of factors that place girls and women "at risk" of entering prostitution and vulnerable to physical and psychological harm once they become engaged in it; descriptions of the vast array of institutional apparatuses, from medicine to law, that have been and continue to be used to control those engaged in the exchange of sex for some valued commodity; and the newfound ways in which some prostitutes and their advocates have resisted and responded to the systems that repress and exploit the former, including undertaking and sustaining an international prostitutes' rights movement.

Consistent with the intent of the book, which is "to show some general comparative trends, directions, profiles, and problems in the social control of prostitutes," readers will come away convinced that a number of things about prostitution transcend cultural context and historical moment. One is that prostitution is overwhelmingly an urban phenomenon. Another is that in every country there is a hierarchy among prostitutes, which is delineated by such factors as location, type of client, level of protection, earnings, and visibility. Finally, and most notably, prostitution is intimately and inevitably connected to the political economy of the society in which it exists and to the system of age, gender, race, and class stratification that characterizes the society in which it takes form.

With regard to gender stratification in particular, the international focus of this book underscores three things. First, to study prostitution is to study the status of women vis-à-vis men. As an Italian prostitute notes in Guido Blumir and Agnes Sauvage's *Done di vita — vita di donne* (1980), "A prostitute is not only a prostitute; she is, above all, a woman." Second, "the selective, gender-biased, hierarchical nature of control often exploits and damages women." And third, the criminalization or legalization of prostitution alone does not stop prostitution; indeed, in her introduction Davis notes that "the more attention given to prostitution as a 'significant' and 'special' social problem, requiring extraordinary measures, the less we see the ordinariness of prostitution as a normal response to gender distinctions and market and political inequalities." This observation is underscored by Finstad and Høigård in the conclusion to their chapter on Norway: "The question of prostitution reaches to the very core of our social organization. If prostitution is to be abolished, we must have a society in which women are no longer forced or coerced to use their sexuality as an economic resource, and in which men are not expected to exploit women's

bodies. The gender-based division of labor is an integral part of society's political economy."

If there is any fault with this book, it has to do with omissions. Key themes and lines of research were not sufficiently developed. For example, the authors devote only minimal attention to the ways in which AIDS has interfaced with, and some would say reorganized, prostitution and has led to what one of them terms "prostitutephobia." Similarly, although prostitution is predominantly a female occupation, the social organization of male prostitution deserves more attention than it receives. Finally, and most important, more information on international trafficking, especially among developing countries, would have enhanced the volume. Despite these omissions, however, the book is unquestionably a very valuable resource for academicians, policy makers, and anyone with a general interest in the status of women globally.

Deciding to Be Legal: A Maya Community in Houston.
By Jacqueline Maria Hagan. Temple University Press, 1994. 200 pp. Cloth, $44.95; paper, $14.95.

Reviewer: KIYOSHI IKEDA, *University of Hawaii at Manoa*

In this study Jacqueline Maria Hagan examines the Totonicapan Guatemalan Maya community in Houston, Texas, before and after the enactment and implementation of the Immigration Control and Reform Act (ICRA) of 1986. A participant-observer, she has developed a framework for organizing rich observations and notes on how undocumented workers have responded to becoming legal immigrants in the United States.

Her work is divided into two parts. Part 1 is made up of three chapters, the first of which provides a framework and methodology for research. In it Hagan presents demographic data on the San Pedro Maya for about 1,000 members of the *munincipio* of Totonicapan, and ethnographies for 74 members (32 women and 42 men). A description of these Maya within their original communities follows in chapter 2, along with a discussion of how some pioneers and their families fostered migration to Houston and employment opportunities there. Hagan discusses Juan Xuc's key role in establishing an employment niche for maintenance workers in a supermarket chain and in developing the network of exchanges from Totonicapan to Houston that allowed for circular migration on a stable basis. In chapter 3 the author provides the demographic and economic context within which men and women adapted and explores the role of strong and weak social-network ties in reproducing Mayan culture within work and community circles. She also addresses how "coyotes" and other interested parties promote illegal immigration by sponsoring and supporting migrants who traverse their way into Houston and return from time to time to San Pedro.

Part 2 focuses on the enactment and initial implementation of the Immigration Control and Reform Act. In chapter 4 Hagan discusses sociolegal, administrative, and individual and community responses to its enactment and details how the law is implemented and administered to assist applicants. In chapter 5 she describes how the ICRA affects opportunities among community members (using length of residence in the United States, marital status, residential status in

Houston and Guatemala, and process of integration into the larger Houston community as measures) and how it continues to impact life and work in and around San Pedro.

Hagan's study of how migrants experienced and adapted to life in Houston and San Pedro before and after passage of the ICRA is a major contribution. She does an excellent job explaining how a viable and adaptive immigrant community for newcomers was created and offers a strong analysis of how network ties enabled men to obtain employment within a carefully cultivated, self-regulating environment. She also does well exploring both the relative absence of weak ties outside the Mayan community for women who must turn to domestic work and the role of women in reproducing Mayan culture within the constraints of Houston. The limitations placed on women's physical and social mobility in fostering networks separate from those of men is well documented, and Hagan's investigation into the differential effects of opportunities associated with becoming legal immigrants, as measured by gender, marital status, length of residence, and residential locus and status, represents a useful application of social network theory and legal impact studies.

Anyone undertaking comparative studies of the ways individuals and communities adapt under conditions of legal and illegal migration and the role of third parties in mediating the migration process will benefit greatly from this work. Given the mix of demographic, sociolegal, and migration-process influences over time, Hagan's effort to weave into her study ethnographic reports of individual migrants is valiant. She might have improved her narrative, however, had she chosen to follow closely the stories of some key figures and families and provided only supplemental reports from the others in her attempt to index the range and types of adaptations the immigrants made.

Citizenship and Nationhood in France and Germany.
By Rogers Brubaker. Cambridge University Press, 1992. 270 pp. $35.00.

Reviewer: CRISTINA ESCOBAR, *San Diego*

This book is a fundamental contribution to the study of migration and citizenship in France and Germany as well as to the analysis of modern state and citizenship in general. It is also as an excellent example of comparative historical methodology. As such, this book appeals to a variety of academic audiences.

Initially, Brubaker helps us see the international arena not just as a territorial organization of states, but also as a set of bounded and mutually exclusive citizenry. This second dimension, he says, is commonly obscured by a territorial bias in the definition of the modern state. Then, he goes into the discussion of modern citizenship. He focuses on the formal dimension of citizenship, which in political sociology has been, more often than not, the taken-for-granted institutional framework. Brubaker's significant contribution is to have brought the modern state in its dual dimensions as a territorial organization and as a political organization into the analysis of modern citizenship. Correspondingly, he proposes two ways of understanding citizenship. One is a formal and abstract approach to citizenship as a general membership status and obligation of a permanent resident population of a state. The other is a concrete and substantive

approach that defines citizenship as special membership status where citizens are a privileged subgroup of the population of the state.

Brubaker unravels the different dimensions of citizenship in France and Germany by means of a historical comparative methodology. The distinction between national and state membership might seem artificial in France, where citizenship as a general status (developed by the absolutist territorial rulers) and citizenship as a privileged status (derived from the classical, medieval, and early modern city-state) were joined in the French Revolution and developed together. However, this distinction is crucial for understanding the immigration problems that France faces today. The French assimilationist and expansive tradition of citizenship has created the problem of second and third generation immigrants who have received citizenship but who have not really been assimilated into the French nation. The comparison with Germany is highly revealing. German nationhood, the privileged status, developed previous to and independently from the state as a result of the particular historical features of the political geography of the region. The German differentialist and ethnic-centered tradition of citizenship policy possesses the paradox of second and third generation immigrants unable and mostly unwilling to achieve German citizenship.

This book is also particularly appealing in its effort to show the significance of the cultural dimension of the politics involved in the definition of citizenship. Even if legislation has been contested, and in some cases modified, Brubaker demonstrates that there are clear continuities, or traditions, in the French and German definitions of citizenship. He argues that these continuities reflect the deep-rooted nature of national self-understanding. He also emphasizes that the traditions are embedded within nonneutral cultural idioms that reinforce and recreate them, producing particular policy outcomes.

Since the author emphasizes that France and Germany are both the exception in Europe, it would have been helpful to have a discussion of the main characteristics of the general context from which they diverge. Nevertheless, Brubaker's contribution to the historical analysis of citizenship in Europe is outstanding.

This book is essential reading for all those involved in the current debate over the future of the nation-state. Whereas others pay attention to the changing characteristics of the international order, Brubaker focuses on deep-rooted traditions. His work makes plain that any reconsideration of the nature and future of the nation-state requires a look at the politics of identity which has played such an important role in the physiognomy of modern Europe.

CHANGING YOUR ADDRESS?

In order to receive your journal without interruption, please complete this change-of-address form and forward it to UNC Press, 60 days in advance, if possible.

Old Address: (Please print)

Name ______________________________
Street ______________________________
City______________________________
State (Country) ________________________
Zip Code____________________________

New Address: (Please print)

Name ______________________________
Street ______________________________
City______________________________
State (Country) ________________________
Zip Code____________________________

Date new address effective:__________________

Social Forces Subscriptions
The University of North Carolina Press
Box 2288
Chapel Hill, NC 27515

CRIMINOLOGY

AN INTERDISCIPLINARY JOURNAL

THE OFFICIAL PUBLICATION OF THE
AMERICAN SOCIETY OF CRIMINOLOGY

EDITOR: Charles R. Tittle
Washington State University

EDITORIAL POLICY: The journal is interdisciplinary, devoted to the study of crime, deviant behavior, and related phenomena, as found in the social and behavioral sciences and in the fields of law, criminal justice, and history. The major emphases are theory, research, historical issues, policy evaluation, and current controversies concerning crime, law, and justice.

Subscription/Membership
Send $50 (for one year) to:

American Society of Criminology
1314 Kinnear Road
Columbus, OH 43212
(614) 292-9207

Manuscript Submission
Send $10 (and four copies) to:

Department of Sociology
Washington State University
Pullman, WA 99164-4020
(509) 334-4003

Index to Volume 74

ARTICLES

BOOK REVIEWS

Jacobson, David (ed.). *Old Nations, New World: Conceptions of World Order.* Alvin Y. So 74(1):348.

Kreps, Gary A., and Susan Lovegren Bosworth, with Jennifer A. Mooney, Stephen T. Russell, and Kristen A. Myers. *Organizing, Role Enactment, and Disaster: A Structural Theory.* Robert A. Stallings 74(2):736.

Kubat, Daniel (ed.). *The Politics of Migration Policies: Settlement and Integration — The First World into the 1990s.* Otis L. Graham Jr. 74(2):735.

Kunda, Gideon. *Engineering Culture: Control and Commitment in a High-Tech Corporation.* Nancy Ditomaso 74(1):369.

Lamont, Michèle Lamont, Marcel Fournier. (eds.). *Cultivating Differences: Symbolic Boundaries and the Making of Inequality.* Anne Bowler 74(4):1441.

Larsen, Otto N. *Voicing Social Concern.* Karen A. Cerulo 74(4):1446.

Lehman, Edward C., Jr. *Women in Ministry: Receptivity and Resistance.* Kevin J. Christiano 74(4):1457.

Lopata, Helena Znaniecka. *Circles and Settings: Role Changes of American Women.* Jennifer L. Pierce 74(2):753.

Lyson, Thomas A., William W. Falk (eds.). *Forgotten Places: Uneven Development in Rural America.* Ronald C.Wimberley 74(1):351.

Miethe, Terance D., Robert F. Meier. *Crime and Its Social Context: Toward an Integrated Theory of Offenders, Victims, and Situations.* Robert M. O'brien 74(4):1449.

Milner, Murray Jr. *Status and Sacredness: A General Theory of Status Relations and an Analysis of Indian Culture.* Bam Dev Sharda 74(1):366.

Monti, Daniel J. *Wannabe: Gangs in Suburbs and Schools.* Malcolm Klein 74(1):358.

Nock, Steven L. *The Costs of Privacy: Surveillance and Reputation in America.* Karen A. Cerulo. 74(1):347.

Obermiller, Phillip J., William W. Philliber (eds.). *Appalachia in an International Context: Cross-National Comparisons of Developing Regions.* Ed Knipe 74(1):352.

Perrucci, Robert. *Japanese Auto Transplants in the Heartland: Corporatism and Community.* Choong Soon Kim 74(2):733.

Poston, Dudley L. Jr., David Yaukey (eds.). *The Population of Modern China.* Ellen Efron Pimentel 74(3):1125.

Reed, Wornie (ed.). *African Americans: Essential Perspectives.* 74(2):746.

Roof, Wade Clark, Jackson W. Carroll, David A. Roozen. (eds.). *The Post-War Generation and Establishment Religion: Cross-Cultural Perspectives.* Kevin J. Christiano 74(1):363.

Sampson, Robert J., John H. Laub. *Crime in the Making: Pathways and Turning Points through Life.* Roland Chilton 74(1):357.

Schneider, Barbara, James S. Coleman. (eds.). *Parents, Their Children, and Schools.* Sophia Catsambis 74(2):751.

Schneirov, Matthew. *The Dream of a New Social Order: Popular Magazines in America, 1893-1914.* Karen A. Cerulo 74(4):1442.

Segal, David R. Segal, Mady Wechsler Segal. *Peacekeepers and Their Wives: American Participation in the Multinational Force and Observers.* James Burk 74(4):1448.

Shields, Rob. *Places on the Margin.* Robert E. Parker 74(2):734.

Simon, Rita J., James J. Fyfe. (eds.). *Editors as Gatekeepers: Getting Published in the Social Sciences.* Michael Schwalbe 74(2):742.

Squires, Gregory D. (ed.). *From Redlining to Reinvestment: Community Responses to Urban Disinvestment.* Nancy A. Denton 74(3):1123.

Strauss, Anselm. *Continual Permutations of Action.* Gary Alan Fine 74(2):741.

Swatos, William H., Jr. (ed.). *Politics and Religion in Central and Eastern*

Clearinghouse on Rural Education and Small Schools

How Schools Can Engage Hispanic Communities

In a new book published by the ERIC Clearinghouse on Rural Schools, Andrea B. Bermúdez of the University of Houston explains reasons and strategies for involving Mexican American and other Hispanic parents in the schooling of their children.

In *Doing Our Homework: How Schools Can Engage Hispanic Communities,* Bermúdez begins by describing the evolution of parents' role in American educatiõn since the century — a time when the country was experiencing a large influx of immigrants. Now, a century later, we continue to face the challenge of addressing the needs of recent immigrants.

Bermúdez provides guidance for educators based on her own research and the research and practice of others. She explains our changing understanding of the role of parents and the persistent barriers to involvement that schools must work to overcome. Strategies include training programs for parents and teachers, model programs for home-school-community partnerships, and collaborations between institutions of higher education and school systems. As Bermúdez points out, in too many places, classrooms have become battlegrounds of competing value systems, learning styles, languages, and attitudes about ethnicity. The casualties have been Hispanic and other minority students who leave school before graduating.

In response, American educators have begun a search that has led some to discover the powerful role parents — including parents who speak little or no English — can play in advancing educational goals. Bermúdez shares their discoveries in this concise (only 91 pages) new resource.

SOCIAL FORCES

University of North Carolina Press
P.O. Box 2288
Chapel Hill, NC 27515-2288

Second class postage paid
at Chapel Hill, N.C.
and additional post offices

RETURN POSTAGE GUARANTEED